Reinforcing the AI4EU Platform by Advancing Earth Observation Intelligence, Innovation and Adoption

# D7.3: Sustainability plan II (final)

| | | | |
|---|---|---|---|
| **Grant Agreement ID** | 101016798 | **Acronym** | AI4COPERNICUS |
| **Project Title** | Reinforcing the AI4EU Platform by Advancing Earth Observation Intelligence, Innovation and Adoption | | |
| **Start Date** | 01/01/2021 | **Duration** | 36 Months |
| **Project URL** | https://ai4copernicus-project.eu/ | | |
| **Contractual due date** | 31/12/2023 | **Actual submission date** | 28/12/2023 |
| **Nature** | R = Document, report | **Dissemination Level** | PU = Public |
| **Author(s)** | Philippe Fournand (BLS), Antonis Troumpoukis (NCSR-D), Xenia Ziouvelou (NCSR-D), | | |
| **Contributor(s)** | Michele Lazzarini (Satcen), Omar Barrilero (Satcen), Manoli Koubarakis (UOA), Richard Hall (Equinor), Andreina Chietera (THA), Stephanie Carpentier (TAS), Mohanad Albughdadi, Jacek Tokarski (CF), Monika Krzyżanowska CF, Giulio Weikman (UNITN) | | |
| **Reviewer(s)** | Richard Hall Equinor, Paula Saameno SATCEN | | |

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 101016798.

## Document Revision History *(including peer reviewing & quality control)*

| Version | Date | Changes | Contributor(s) |
|---|---|---|---|
| V7.3 | 2023-09-19 | Initial draft | Philippe Fournand BLS |
| V 7.3.1 | 2023-10-26 | Second draft | Philippe Fournand BLS |
| V.7.3.2 | 2023-11-28 | PRe-final version | Philippe Fournand BLS |
| V7.3.3 | 2023-12-04 | CLosed version integrating contributions from partners | Philippe Fournand BLS |
| V7.3.4 | 2023-12-15 | Final version for review | Philippe Fournand (BLS) Xenia Ziouvelou (DEM) |
| V7.3.5 | 2023-12-22 | Final Version after review | Philippe Fournand (BLS) |

## Executive Summary

This deliverable describes the AI4Copernicus approach to sustainability to ensure the further uptake of the project's main innovation and technology achievements. The report presents in section two **the Key Exploitable Results (KER) of the project**, developed by the partners and how they intend to use and exploit them. In **section three,** the same approach is set to present the exploitation strategy carried out by the open call winners through the design, test validation and deployment of **thematic services**. Finally, the document proposes in **section four** a **pathway to develop an AI/EO (Earth Observation) node** in partnership with the AI On-Demand Platform (AIODP https://www.ai4europe.eu/), highlighting the challenges faced by Europe in that area.

**AI4Copernicus delivered exciting results for all partners** and the AIODP.

**42 assets** were developed: **17 Bootstrapping Services (BR)** and **25 Thematic Services (TS) developed by the Open Calls Winners**

The technological developments made through **the bootstrapping services** allowed traditional EO players (SATCEN, EMCWF, CF), research organisations (DEM, UNITN) or companies (Equinor, THA) to **develop or enhance tools or models in real conditions with open-call winners**. The AIODP and their platforms will further maintain and develop these results.

The project has been, for all partners, **a steppingstone to understanding better processes and pathways to build a more user-centric EO**, solving challenges and concrete problems, thus mainstreaming the usage of EO data and triggering a wave of AI services powered by EO Data. Public and Private companies have strengthened their market intelligence and tested open innovation pathways that will be further expanded.

The consortium built **common tools**: **an educational module, a GitHub, a methodology to enhance the product market fit for open call winners,** involving end-users in the consortium, creating a Trustmark and helping companies to pitch their ideas. It's paving the way for a methodology aiming to accelerate growth for start-ups from the space industry, which echoes very well with the ESA practices.

**Fruitful collaborations** have been tested and will be pursued between partners, such as the Earth QnA Engine or between some partners and open-call winners. The 27 projects awarded have developed AI assets powered by EO Data on critical challenges. If the company's capability to grow is difficult to foresee, the consortium remains confident and has deployed tools to support their growth.

Finally, we use feedback from partners and the open call winners, seconded by an extensive European and international benchmark**, to propose several scenarios for building an AI/EO node** for the next phase of the AIODP. In a context of fierce global competition, the consortium firmly believes in the added value of this node to significantly **contribute to a more AI/EO user-centric approach**. Indeed, AI needs data, and EO can bring massive amounts. **AI applications powered by EO data** can **solve critical challenges** such as climate change, security or precise agriculture. The research agenda to reach a user-centric EO exists in several strategic domains highlighted by the report, such as **foundational models**. We designed several pathways for creating this AI/EO node, with different services or packages of services adapted to different targets. The consortium believes the AIODP and the subsequent follow-up project, Deploy AI, can build a strategy upon these elements.

## Table of Contents

## List of Figures



## List of Terms & Abbreviations

| Abbreviation | Definition |
|---|---|
| **AIODP** | AI On-Demand Platform |
| **AI** | Artificial Intelligence |
| **BR** | Bootstrapping Services |
| **DIAS** | Data and Information Access Services |
| **EO** | Earth Observation |
| **ESA** | European Space Agency |
| **KER** | Key Exploitable Results |
| **MVP** | Minimum Viable Product |
| **NASA** | National Aeronautics and Space Administration |
| **OC** | Open Calls |
| **TS** | Thematic Services |

# 1 Introduction

## 1.1 Purpose and Scope

The deliverable presents the final AI4Copernicus sustainability plan. The AI4Copernicus project has bold ambitions: **bridging the gap between the EO community and the European AI community, with a focus on the stakeholders gathered around the** AIODP. Some well-informed experts point out that the EO community and system are now reaching a tipping point, a sort of "iPhone moment" [1]that will have enormous impacts. Thus, bridging the gap between EO and AI is of the utmost importance in the long run.

The AI4Copernicus sustainability plan encompasses the following dimensions:

- **Section 2 Sustainability of the Solutions/Services proposed by the consortium (BR, General Purpose Semantic Services and other results.** See also D6.5 for an overview of the services**).** How will the services produced by the AI4Copernicus project be further maintained and expanded after the end of the project? Section 2 presents the individual strategy proposed by each partner.

- **Section 3 Sustainability of the TS and the solutions developed by the project's third parties, the consortium selected in the five open calls implemented by AI4Copernicus.** How will the solutions designed by the AI4Copernicus early adopters, the open calls awarded consortium, could be deployed, maintained and scaled after the end of the project? What are the lessons learned in terms of supporting Starts-Ups and Smes?

- **Section 4 Ecosystem Sustainability and Community deployment**. What will happen after the end of the project? How could the community be convinced to use AI4Copernicus services on the AI On-Demand Platform in the long run? What proposals does the consortium, beyond the individual exploitation plans, issue to capitalise on AI4Copernicus results and reach a new development cycle through the AIODP? Is there a business model of potential funding to further expand AI4Copernicus? The AI/EO node concept is particularly relevant in offering tangible community perspectives.

## 1.2 Approach for Work Package and Relation to other Work Packages and Deliverables

The deliverable is a consortium collective outcome linked with all WPs, particularly WP2 (assessing the BR and TS) and WP6, presenting the open call results.

## 1.3 Methodology and Structure of the Deliverable

The methodology adopted to build the deliverable is the following

[1] Aravind Terra watch From GHG to ESG: Demystifying Earth Observation for Climate (substack.com)

- Creation of an exploitation Working Group gathering one representative per partner
- Administration of two questionnaires to all partners. The first is a Google doc that concerns their exploitation strategy related to their **KERs**. The second questionnaire examines their vision to develop an AI/EO Node starting after the end of the AI4Copernicus project.
- Partner BLS conducted a deep AI/EO ecosystem analysis, seconded by reviewing key trends and roadmaps such as FIRE in the domain.
- A seminar with the Exploitation Working Group took place on 21 April 2023 to discuss the outcomes and further develop a sustainability strategy, gravitating around the AI/EO node concept.
- Athens Ecosystem Forum offered the opportunity to collect feedback and expectations from the open-call winners during a world cafe session organised around three main topics:
    - Access and use of EO Data.
    - The added value of AI4Copernicus SERVICES to mainstream EO and AI usage.
    - A future? AIOD EXISTING & NEW SERVICES & AI/EO node.
- A document highlighting the main conclusions has been drafted, and the findings have been reinjected in this document.
- The document was extensively discussed between partners to reach the final document.

The structure of the deliverable is as follows: **Section 2** provides an overview of the AI4 Copernicus partner's exploitation strategy. **Section 3** presents the Open Call winner strategy (how?) **Section** 4 details the challenges of deploying an AI/EO node and proposes a pathway to address them.

# 2 AI4 Copernicus exploitation plan

This section presents the exploitation plan of the KERs for the AI4Copernicus partners: BR, General purpose Semantic Services and other results.

## 2.1 AI4Copernicus vision and overarching vision

Earth Observation is living its "iPhone moment"! Observing the Earth from the sky is not something new. It backs from the first satellites launched and is a domain widely invested by governmental agencies (ESA, NASA). However, EO applications and domains have exploded with the exponential progress made by the technologies during the last decade and the dramatic cost decline. As a result, new business models have dramatically changed the landscape.

Experts now foresee a trillion-dollar market in 2040[2] organised around a value chain involving satellites (manufacturing, launch, exploitation), broadband and Internet, analytics, AI, and main EO providers, becoming a commodity for many industrial players, large and small companies. Competition between superpowers and now super tycoons is the new rule. Europe's sovereignty in this competition is at stake, not only for providing data at affordable costs but also to exploit EO and develop new industrial applications to mitigate and adapt to climate change, manage risks, and develop precise and frugal agriculture, to name a few of the challenges addressable by using EO combined to AI.

Thus, the vision brought by AI4Copernicus and shared with other similar or sister projects (ICT-49 projects) is particularly important to Europe. **Enabling and supporting EO data-driven AI Innovation and Services in Europe,** particularly in 4 of the most important industries[3] (agriculture, energy, security, and health), may lay the foundation for more stable cooperation between two key European communities:

- **AI community**: represented by the AIODPand the
- **EO community** brought by AI4Copernicus.

Deliverable 7.2 (Sustainability Plan Interim Report M18) provided a more detailed analysis of the market, trends, and vision of AI4 Copernicus.

[2] Morgan Stanley Research Space: Investing in the Final Frontier July 24 2020. Strictly speaking, the total market addresseable at 2040 is broader than the market address by AI4Copernicus is data, analytics and insights extracted from EO data (see next pages)

According to the 2019 PWC Copernicus Market Report (PwC, 2019) in the 2008-2020 period, Copernicus data created economic value in the range of 16 to 21 billion Euro, in addition to other, possibly larger, non-economic impacts (e.g., related to environmental factors, security, and other social impact for example). Similarly, the latest Geospatial Industry Outlook and Readiness Index- GeoBuiz, indicates that the Earth observation industry as a whole is estimated to be worth almost US $58 billion in 2019, rising to almost US $76 billion in 2020.

[3] AI4Copernicus has, by design, selected to focus on some of the most impacted areas, namely *energy, agriculture, security, and health* related ones, which according to the PwC research (Copernicus Market Report, 2019) covered the bulk of the economic impact of the Copernicus programme during the past decade (i.e., for more than 80% of the economic benefits for end users). These four primary AI4Copernicus focus areas have also been some of the fastest growing ones - with double-digit average annual growth, more than 20% for agriculture and (renewable) energy for example.

## 2.2 AI4Copernicus Services and assets

### 2.2.1 AI4 Copernicus Service Portfolio

In three years, the AI4 Copernicus consortium developed a portfolio of services aiming to connect EO and the AI world.

| Solve user's problems through application and innovation challenges | Access to data through DIASES and train data | Find usable data General Purpose Semantic Service | Bootstrapping Pre processing services | Bootstrapping Use realistic ML Models |
|---|---|---|---|---|
| **User's stories Energy**<br><br>**Open calls** addressing challenges in four areas energy, security, health and agriculture<br><br>**Mentoring services to access and use data**<br><br>Ethical Support, trustmark and support to pitch solutions | **CREODIAS Access**<br>**WEKEO Access**<br><br>Cloud and computing services<br><br>**CREODIAS Deployer** from AI4 Copernicus to **access AI4 Experiment** | **Semantic Search and tools**<br>- GeoTriples<br>- JED ai Spatial<br>- Semagrow<br>- Sextant<br>**Earth Observation QnA** | **Satcen (Security)**<br>**Sentinel-1 GRD** pre-processing<br>**Sentinel-2** pre-processing<br>**Sentinel-1 Change detection–** Amplitude Change Detection and Multi-temporal Coherence<br>**Sentinel-2** Change detection<br><br>**UNITN (Agriculture)**<br>TimeSen2Crop<br>Harmonization of pre-processed Time Series of Sentinel-2 data<br>**ECMWF (Health)**<br>Probabilistic downscaling of CAMS air quality model data | **THA** (TimeSen2Crop)<br>**Deep network** for pixel-level classification of S2 patches (agriculture)<br><br>**UNITN (Agriculture)**<br>**Long Short-Term Memory Neural Network** for NDVI prediction<br>**Long Short-Term Memory Neural Network** for Sentinel-2 for crop type classification<br>Pre-Trained Long Short-Term Memory for crop type classification |

*Figure 1 AI4 Copernicus Service Portfolio*

The AI4Copernicus project brings **five categories of KERs for enriching the AIODP**, as shown in the table above.

The first category is often neglected but is one of the most important. AI4 Copernicus's outstanding **open-call activities** allowed 27[4] projects to test and develop AI solutions using EO data and user stories. User stories are real-world problems faced by end-users. The mentoring services were extended **to ethical support, the creation of a Trustmark to follow the maturity of the solutions on four standpoints** (deliverable technical, ethics, business performances see D 6.5) **and the relations with Venture Capitals** (through the Vc booklet and the website see D6.5)

The second range of services is **accessing data** through CREODIAS and WEKEO and providing cloud and computing services.

The third one enables users to **find usable data** through General Purpose Semantic Services, particularly the Earth QnA and the semantic search and tools. Those services have been tested by two projects of the 5th open call.

The fourth and fifth are bootstrapping services offering preprocessing services or Machine Learning Models.

**The whole package of AI4 Copernicus services shows a first attempt to deliver a user-centric AI/EO approach**.

The table below illustrates the number of projects using each service. In particular, for the 1st, 3rd, and 4th Open Calls, we show the number of projects using a service in the pipeline of their final product and, in addition, the number of (additional) projects that have conducted some small-scale

[4] AI4Copernicus selected 29 projects.27 developed TS, the 2 remaining ones are the citizen use cases selected in the OC 2Open Calls for social challenges and used for OC4 (Open Call for use cases based on social challenges)

or large-scale experiments but didn't eventually use their service in their final product. Regarding the 5th Open Call, since its topic was about testing the AI4Copernicus services and not producing new products, for each service, we illustrate the number of projects that expressed their interest in trying the specific service vs. the number of projects that tested the service in their final program.

| Service | number of projects using the service | | |
|---|---|---|---|
| | 1st Open Call | 3rd/4th Open Call | 5th Open Call |
| **Security Services KERs (SatCen)** | used in final product: **2** used in experiments: **1** | used in final product: **2** used in experiments: **1** | expressed intention: **6** used in experiments: **4** |
| **Agriculture Services KERs (UNITN)** | used in final product: **1** used in experiments: **1** | used in final product: **2** used in experiments: **2** | expressed intention: **4** used in experiments: **6** |
| **Agriculture Services KER (Thales)** | used in final product: **0** used in experiments: **2** | used in final product: **1** used in experiments: **1** | expressed intention: **7** used in experiments: **5** |
| **Health Services KER (ECMWF)** | used in final product: **0** used in experiments: **1** | used in final product: **0** used in experiments: **1** | expressed intention: **2** used in experiments: **3** |
| **General Purpose Semantic Service KER (UoA)** | used in final product: **0** used in experiments: **0** | used in final product: **0** used in experiments: **1** | expressed intention: **3** used in experiments: **2** |

*Figure 1 Use of AI4 Copernicus service by open call winners*

For a detailed analysis of the use of bootstrapping services, see D2.3: Validation Plan and Evaluation Report and D6.5 Service extensions and optimisations & policy guidelines for AI adoption. This latest deliverable reports the positive feedback from 5th OC winners when using the AI4 Copernicus services.

### 2.2.2 AI4 Copernicus exploitation plan

This section covers the partner's exploitation plan, presenting the assets they developed, their description, the TRL reached and the link to the AIODP.
In total, AI4 Copernicus developed 17 Bootstrapping services

- 6 in the Security domain
- 5 in the Agriculture domain
- 1 in the Health domain
- 5 General Purpose Service.

For more information, see Deliverable 2.4 for an extensive analysis of their usage.

### *2.2.2.1 Overview of the Bootstrapping services and assets developed by AI4 Copernicus partners*

The table below presents the services and assets the AI4 Copernicus partners developed and the link established with the AIOD.

*Figure 2 Bootstrapping services and assets developed within AI4 Copernicus and their integration in the AIODP.*

| Partners | Services | Description | TRL end of the project | LINK AIODP |
|---|---|---|---|---|
| CF<br>Cloud Ferro<br>KER Cloud and computing services<br>CREODIAS Deployer | AI on Demand Platform integration – link Acumos and Creodias PaaS (K8S) Allowing users to deploy K8S clusters with their AIoD solutions on the CREODIAS platform. | Generic Services<br>• Data provision,<br>• Computing,<br>• Infrastructures,<br>• Testing,<br>• Marketplace and Security<br>Dedicated Services<br>• Docker registry (GoHarbor)<br>• Dedicated K8S | Docker registry (GoHarbor) – TRL 7<br><br>Dedicated K8S – TRL 6<br><br>AIonDemand Platform integration – TRL 4 | Fully Achieved |
| University of Athens<br>UOA<br>KER General Purpose Semantic Service | EarthQnA<br><br>Copernicus Ontology:<br><br>Linked Geospatial Data Tools | A question-answering engine for the discovery of satellite images. The engine will run on top of CREODIAS.<br><br>An ontology for Earth Observation, and the Copernicus program and its various components.<br><br>Tools for transformation (GeoTriples), interlinking (JedAI), storage/query (Strabon) and visualisation (Sextant) of linked geospatial data. | TRL 4 Demonstration with real CREODIAS metadata<br><br>Tools for transformation (Geo Triples) interlinking (JedAI) storage/query (STRABON) and visualisation (Sextant) of linked spatial data | Through CREODIAS Interaction to develop with AIODP search engine.<br><br>Copernicus Ontology will be part of the AIODP ontology.<br><br>The tools are available in the AIODP. |

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 101016798.

| | | | | |
|---|---|---|---|---|
| Partners | Services | Description | TRL end of the project | LINK AIODP |
| THALES (THA)<br>KER BR Agriculture | Deep Network for pixel-level classification of S2 patches | Users can train and test the service using the SEN12MS dataset or provide their data under the Sentinel-2 format to train a new segmentation model. A various number of parameters are available to customise the training phase. This asset also provides the code to apply the already trained model on SEN12MS to a new set of data. | Fully operational<br><br>TRL 7 and 8 | AIODP Catalogue |
| SATCEN<br>Ker BR Security | Set of Security Bootstrapping services | Reducing the time and resources of the bidders in the data preparation allows them to focus on developing innovative services based on AI. They include pipelines for the preprocessing and change detection of Sentinel-1 and Sentinel-2 imagery.<br>The modules include S1-preprocessing, S2-preprocessing, S1 change detection and S2 change detection | Fully operational<br><br>TRL 7 and 8 | AIODP Catalogue |
| ECMWF<br>KER Health BR | Probabilistic downscaling (super-resolution) of CAMS air quality (AQ) and atmospheric composition (AC) model output | This service super-resolves air quality and atmospheric composition to a better spatial resolution to identify pollution or greenhouse gas emission hotspots.<br>The service can be used in a jupyter-lab environment or using command-line | Fully operational<br>TRL 8 | AIODP Catalogue |

| | | | | |
|---|---|---|---|---|
| Partners | Services | Description | TRL end of the project | LINK AIODP |
| UNITN<br>KER BR Agriculture | AI4 Copernicus Agriculture Bootstrapping services | Two neural networks with different targets: one for classifying crop types, the other for predicting indices (LST memory neural network for Sentinel2, pre-trained LST Memory)<br>A harmonisation tool to standardise the input images is available to the user, allowing a preprocessing of the raw S2 acquisitions.<br>An Open-Source Dataset, TimeSen2Crop, can be downloaded to train the model using a million already extracted training samples. | TRL 8 | AIODP Catalogue |
| EQUINOR<br>KER: Market insight | User stories Energy | Used for the open calls energy precisely defining user's needs | Not relevant | AI4 Copernicus user stories on the AIODP<br>Six users' stories are accessible on the AIODP |
| NCSRD (DEM)<br>KER General Purpose Semantic Service | Semagrow Federated Query Engine | Semagrow is a SPARQL query federator of heterogeneous data sources. It allows combining, cross-indexing and generally making the best of all public data, regardless of their size, update rate, and schema. Semagrow offers a single SPARQL endpoint that serves data from remote data sources and that hides from client applications heterogeneity in both form (federating non-SPARQL endpoints) and | TRL8 | The tool is available on the AIODP Catalogue. |

| | | | | |
|---|---|---|---|---|
| | | meaning (transparently mapping queries and query results between vocabularies) | | |

TAS, INSEAD and BLS were focused on other tasks and didn't develop bootstrapping services. Their KERs are presented in the next section.

### 2.2.3 Exploitation plan at a glance

The exploitation strategies are different from the partners.

- Public institutions
- Companies
- Research organisations

#### *2.2.3.1 Public institutions.*

Public institutions (Satcen/ECMWF) expect to channel more users or consumers on their platforms, data and services. AI4Copernicus offered the possibility to test the deployment of services, enhancing the added value of their platforms (e.g. ECMWF) or the portability of their tools (e.g. SATCEN) and to follow (through the open calls) innovative projects that can help them support their needs. However, being non-profit entities working in the public domain, they have not defined any business plans to go to the market, and they will exploit the services internally to enhance their operational capabilities.

In any case, the assets and services are included in their platform and the AIODP, contributing to enriching their services.

- SATCEN is the partner in charge of the Security Domain, particularly **interested in the process and the results of the open calls**. Regarding the process, **the objective is to learn by doing how to improve applications and enlarge** SATCEN **capabilities.** The approach and the way of working from the open call bears fruitful lessons for SATCEN. Moreover, one of the products developed by one of the open calls projects is a super-resolution algorithm that can increase the quality of the services SATCEN provides. The company **developing the product and SATCEN are exploring the possible use within SATCEN tools.**

#### *2.2.3.2 Companies*

- CF, private company managing access to DIAS and core partner of the Copernicus Data Ecosystem https://dataspace.copernicus.eu/**.** CF's strategy is to **mainstream and increase the number of Copernicus Data Ecosystem Services user**s.

  However, the usage is limited to **people who already know what they want and how to navigate the CREODIAS services**. This is the objective of the seamless integration of CREODIAS services in the AI4 Experiment part, to bring new users from the AI world. The integration could also allow CREODIAS to provide resources to a follow-up project, particularly for creating an AI/EO node on the AIODP platform. **A second strategic goal for CF is to add a powerful search engine to its services** that can dramatically facilitate the search for resources. Today, the search is through categories of data sources (Sentinel 1,

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 101016798.

Sentinel 2, etc.), imagery types (thermal, radar) or topics (agriculture, aeronautics, etc). This structure reinforces the bias towards professional users, as mentioned above. One critical aspect of solving **this issue is the synergies between CF and UOA concerning the Earth Observation QNA engine**. At the time of this report, **discussions have started** between the two partners concerning a potential integration of the EOQnA Engine in the CREODIAS Copernicus Data Ecosystem Services. However, at the moment, only a demonstration on CREODIAS metadata has been achieved.

- **Equinor** is looking for a (or several) platforms of choice to make open innovation, bundling challenges, data access and processing, open calls on specific problems and solutions that can be developed on an open-source mode or more private ones. The rationale is they don't want to invest too much outside of their core business, and they are looking for partners to outsource parts of the open innovation process (this is insight as a service and challenges as a service). The model is like InnoCentive or Kaggle but based on AOI and EO. Equinor, through the open calls, had also identified start-ups that could provide Equinor with services and exposed the company to a wider audience. If Equinor is the most advanced partner, other partners like Satcen have brought particular interest in the open calls to scout solutions, such as super-resolution algorithms that can be hosted on their platform (see above).

  Equinor also took advantage of the open calls to set up user stories that the company will use later to develop research, academics or training partnerships. See below the AI4 Copernicus story for Equinor, summarising the expected outcomes and benefits of AI4 Copernicus for the company.

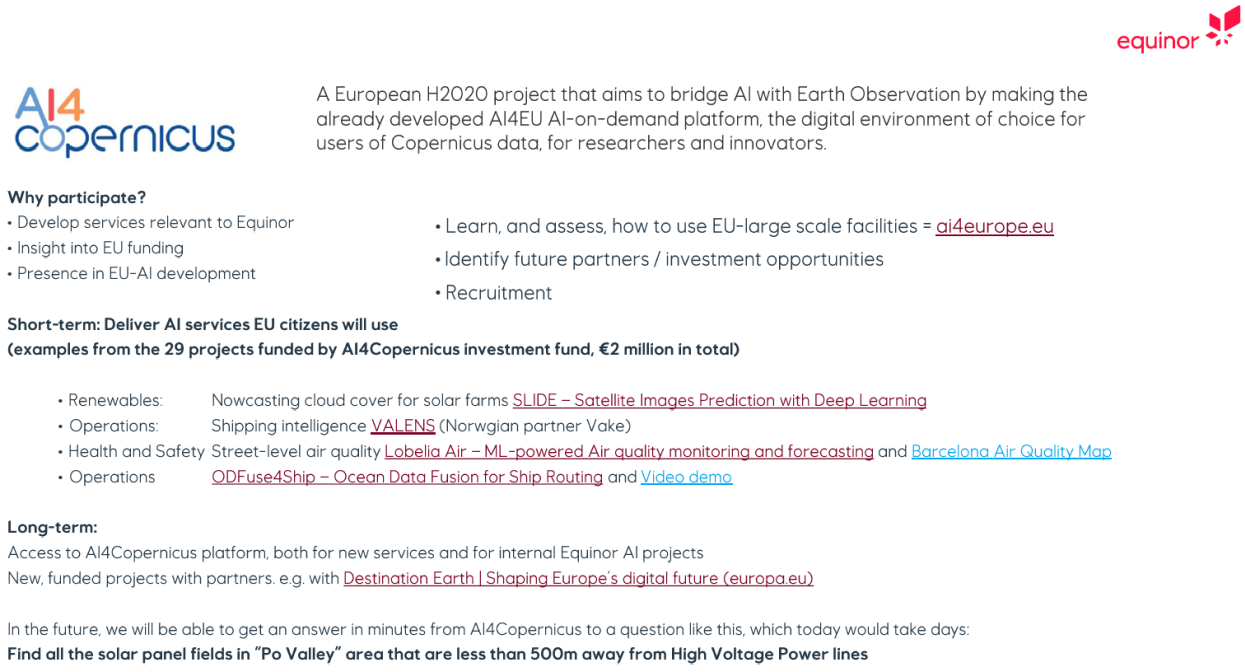



*Figure 3 AI4 Copernicus one page for Equinor*

- **Thales:**

AI4 Copernicus is a crucible for fostering open innovation within THA research labs.

1. **Integration with Thales Products: Maximising Internal Synergies**
   As part of its initial objectives, THA is dedicated to seamlessly integrating AI4 Copernicus services into THA products. This integration will streamline the incorporation of these services into THA's internal business divisions, thereby elevating the synergy between research labs and product development efforts.
2. **Continued Collaboration and Learning:**
   In addition to the strategy mentioned earlier, THA is actively exploring the possibility of establishing new agreements with a specific focus on retrospective analysis (retex) of the bootstrap service (Deep Network for pixel-level classification of S2 patches) provided to the successful consortia in the AI4 Copernicus open calls. The primary aim of this retex initiative is to assess how this service performs and its impact in real-world scenarios. Furthermore, THA intends to leverage the insights gained from these reports when this service is applied to new data or unforeseen applications. This forward-thinking approach underscores our commitment to maintaining adaptable and pertinent offerings that align with the evolving requirements of our partners and clients.
3. **Licensing Strategy: Unlocking Opportunities Beyond**
   THA valuable assets will be available for external utilisation through a licensing strategy. After the AI4 Copernicus project's culmination, THA intends to introduce a Royalty Business model for licensing. This approach is designed to benefit small and medium-sized enterprises (SMEs) and startups keen on integrating THA assets into their operations. By doing so, they gain a competitive edge by leveraging THA technology, reducing their development timeline, and positioning themselves more effectively in the market. However, it's important to note that this endeavour requires careful consideration. It entails a new mindset and establishing new administrative and juridical tools, which may take some time to develop and implement. This axis represents a long-term commitment to exploitation and growth.

AI4 Copernicus is a testimony to THA's dedication to innovation and a steppingstone towards a more expansive and collaborative future. THA looks forward to harnessing these potential exploitation avenues to drive both internal growth and external partnerships while continually enhancing the value of our services through retrospective analysis and adaptability.

#### *2.2.3.3 Research partners*

**Research partners have non-commercial exploitation plans that are linked to research and research community evolution based on the knowledge acquired via AI4Copernicus.** Assets and Services are OSS and can be further improved in new projects.

- UOA, one key exploitable result, is of great interest to mainstream the usage of EO data and services. It can represent a core component of a future AI/EO node: **the EO QnA Engine developed by UOA.** As mentioned above, searching data is complex and a barrier to

mainstream EO datasets and usage. Indeed, extracting meaningful information from EO raw material data is an essential prerequisite to building added-value applications and achieving impact.

*"Making EO data accessible in catalogues is not enough. Improving users' experience when working with it must also be high on the agenda. Such experience is currently still not optimal since EO data analysis and interpretation is still performed in a very laborious way, by repeated cycles of trial and error, without reaching the desired degree of flexibility and robustness. Furthermore, the overwhelming amount of data puts strong limitations on the extent (geographic coverage within acceptable cost bounds), depth (causal analysis, evaluation of the implications) and response time of human interpretation. Consequently, a significant amount of EO data remains unfound and unused. User-centric EO is a concept that enables the interpretation of EO data content, associates it with other sources of information, understands user inquiries in an interactive dialogue (thus refining the expression of needs), distils data content, and suggests the most appropriate algorithms information items or the interpretation alternatives."*[5]

*Similar to what search engines do on the Internet, a Remote Sensing Visual Question Answering (RSVQA) engine could allow anyone (from scientists to laymen and journalists) to retrieve the relevant information contained in the images.*[6]

The Earth QnA Engine aims to make **EO dataset discovery like searching with Google and target non-expert EO data users.** EarthQA accepts questions in natural language (English), asking for EO datasets with specific properties and returning links to such datasets. The properties can refer to satellite image metadata and relevant geographical knowledge from the KG DBpedia. Examples of questions

- Find Sentinel-1 products that show Etna in March 2018.
- Find Sentinel-2 MSI products with cloud cover below 10% during March 2017 / 2018
- Find Sentinel-3A Water Full Resolution (WFR) products with the data collected in January 2018.
- Retrieve all GRD Sentinel-1 images that cover the Black Sea and have been taken from 1/06/19-1/15/19.

The EO QnA Engine represents a State of the Art. QWANT has developed a similar search engine during the project Snap Earth [7] , but it seems not operational at the moment.
The EO QnA Engine is working to access archived data and finalise the demonstration. Once it is done, it can represent one of the core components of the AI/EO node on the AIODP that should also be compared to the expected performance from large language models for Earth Observation models that ESA would soon launch.

[5] Artificial intelligence to advance Earth 15 may 2023 observation: a perspective Devis Tuia, Konrad Schindler, Begum Demir, Gustau Camps-Valls, Xiao Xiang Zhu, Mrinalini Kochupillai, Saso Dzeroski, Jan N. van Rijn, Holger H. Hoos, Fabio Del Frate, Mihai Datcu, Jorge-Arnulfo Quian_e-Ruiz, Volker Markl, Bertrand Le Sau, Roxchelle Schneider
6 IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE, PREPRINT, FULL VERSION: 10.1109/MGRS.2020.3043504 April 2021 Towards a Collective Agenda on AI for Earth Science Data Analysis. Devis Tuia, Senior Member, IEEE, Ribana Roscher, Member, IEEE, Jan Dirk Wegner, Nathan Jacobs, Senior Member, IEEE, Xiao Xiang Zhu, Senior Member, IEEE, Gustau Camps-Valls, Fellow, IEEE
[7] https://snapearth.eu/services

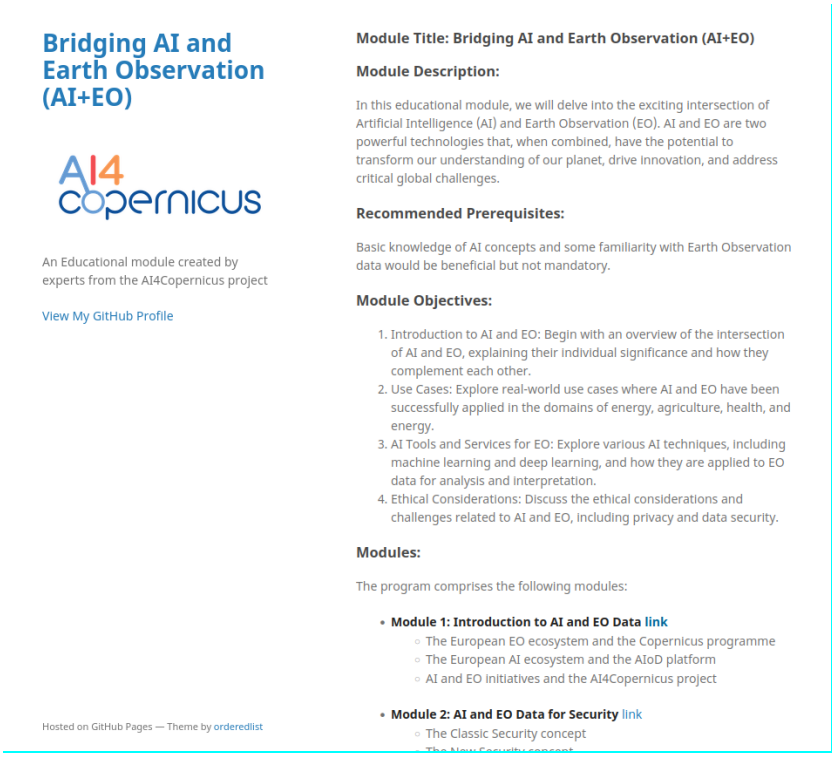
### *2.2.3.4 AIEO educational module*

AI4Copernicus has built an Educational Module titled “Bridging AI and Earth Observation (AI+EO)”. Its purpose is to delve into the exciting intersection of Artificial Intelligence (AI) and Earth Observation (EO) since “AI and EO are two powerful technologies that, when combined, have the potential to transform our understanding of our planet, drive innovation, and address critical global challenges”. This module is designed for students, researchers, start-ups, and anyone interested in harnessing the combined power of AI and Earth Observation to address environmental, societal, and economic challenges. By the end of this module, participants will have a comprehensive understanding of how AI and EO can work together to unlock new insights, drive innovation, and contribute to a more sustainable future. Basic knowledge of AI concepts and some familiarity with Earth Observation data would be beneficial but not mandatory.

In a nutshell, the module has the following content:

- Introduction to AI and EO: Begin with an overview of the intersection of AI and EO, explaining their individual significance and how they complement each other.
- Use Cases: Explore real-world use cases where AI and EO have been successfully applied in the domains of energy, agriculture, health, and energy.
- AI Tools and Services for EO: Explore various AI techniques, including machine learning and deep learning, and how they are applied to EO data for analysis and interpretation.
- Ethical Considerations: Discuss the ethical considerations and challenges of AI and EO, including privacy and data security.

The module can be accessed through its GitHub page at https://ai4copernicus-course.github.io/ (see Figure below). The content is publicly available, and the course format is self-paced. The course modules can be extended to other thematic areas and the existing ones can be updated.



Another educational resource provided by the AI4Copernicus consortium is a half-day tutorial titled "AI4Copernicus tools and methods for bridging AI and EO". This tutorial was organised in the Big Data from Space 2023 – BIDS conference, held from 6 to 9 November 2023, in Vienna, Austria, and was presented on-site by the AI4Copernicus colleagues: Antonis Troumpoukis (NCSR-D), Iraklis Klampanos (NCSR-D), Despina-Athanasia Pantazi (UoA), Omar Barrilero (SatCen), Giulio Weikmann (UNITN), Mohanad Albughdadi (ECMWF) and Vasileios Baousis (ECMWF). In this tutorial, we presented the main technological assets AI4Copernicus brings and its methodology and tools for linking the DIAS and AIOD platforms through appropriately selected use cases during a hands-on session. The hands-on session took place on the CREODIAS infrastructure provided by partner CF and involved the application of AI techniques on available datasets. All the material, presentations, instructions, and documentation are publicly available on the tutorial website at GitHub: https://ai4copernicus-bids2023.github.io/

### 2.2.4 Individual Exploitation Strategy in a nutshell

*Figure 4 AI4 Copernicus Exploitation pathways*

| Partner | Pathway | Exploitation Path |
|---|---|---|
| NCSR-D | Open-Source Exploitation | NCSR-D plans to exploit the development and integration of ML for EO methods to develop novel tools and further its expertise in the AI and EO fields. It will ensure that the AI-EO activities will be successfully sustained in the EO vertical within the AIoD platform and via its participation in the new EU-funded project "DeployAI" that will focus on deploying the AIoD platform.<br>NCSR-D will also ensure that the Trustworthy AI activities and knowledge acquired within the AI4Copernicus will evolve and align with the market demands and EU legal framework within the AIoD platform.<br>In addition, it will also exploit its experience in organising Open Calls and innovation activities to extend its network and innovation capacity. |
| UoA | Open-Source Exploitation | The EARTH QnA engine will be made available as open source, and we expect to reuse it in other future projects where we will be involved. We will also work closely with CF to make the engine available on CREODIAS. We can also use it in other DIAS platforms.<br>ESA is building a digital assistant similar to that with more functions. EarthQA is currently being extended as part of a digital assistant for satellite image archives in ESA project DA4DTE (October 2022-March 2024). |
| TAS | **Know how.**<br>Commercial Exploitation | TAS is mainly involved in WP2, aiming to develop a better understanding of end-user needs and evaluate the services and platforms.<br>The project's outcomes will have no direct commercial exploitation but indirect effects due to a better knowledge of end users and potential connections between AI/EO, TAS, and potential thematic partners. An example of an application could be a partnership for additional services for DestinE Core Service Platform (DESP), a key element of the European Commission's flagship initiative Destination Earth (DestinE) |
| INSEAD | Commercial exploitation – (indirectly from the | Lessons learned from AI4 Copernicus will reinforce the INSEAD AI Trustworthy research and education field. |

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 101016798.

| | | |
|---|---|---|
| | AI4Copernicus know-how & expertise) | |
| Partner | Pathway | Exploitation Path |
| THA | Commercial Exploitation | AI4Copernicus is for THA, a test bed to develop open innovation inside the research labs. Today, they are working for their business units. Tomorrow, part of their activities could be directed towards external customers. Thus<br>THA have two strategies:<br>The asset is available under the licensing strategy. After the end of the AI4Copernicus project, the intent is to deliver the license using a Royalty Business model. The idea is to address SMEs or startups that want to incorporate our asset, enabling them to complete their product using THA technology, shorten the development phase and compete with others in the market.<br>Integrate the services in a THA product and transfer it to internal business lines. |
| ECMWF | Open Source Exploitation<br>&<br>Commercial Exploitation | The ECMWF business model aims to enhance the use of CAMS and C3S data by providing open-source tools to exploit these datasets to provide added-value services (such as high-resolution air quality maps). No economic benefits are expected from the service but to increase the uptake and use of ECMWF data for different vertical markets. To ensure the sustainability of the developed tool, it can be included in the next release of CAMS tools. |
| CF | Open Source Exploitation<br>&<br>Commercial Exploitation | The expected results could lead to three routes for exploitation: **the development of CF services through the new Copernicus Data Ecosystem to new users from the AIODP.** These new customers will result from the federation between the two platforms.<br>The second exploitation mode is **to open routes for a more user centric EO approach through the potential partnership with the EOQnA engine** and the lessons learned through the open calls. The third route could be to **propose services for developing an AI/EO node on the AIODP.**<br>The cluster currently used for integration with AI4Experiments will be decommissioned upon the completion of the project. The solution is designed in such a way that users can utilise their clusters created within the CF infrastructure. After the project, **users can use the Creodias Deployer solution, leveraging their resources on CF clouds.** A tutorial on creating a Kubernetes cluster is available at How to Create a Kubernetes Cluster Using CREODIAS OpenStack Magnum. All necessary information for creating an account in CF and accessing resources all the time will be available at |

| | | |
|---|---|---|
| | | CREODIAS documentation. AI4EU Integration with Creodias Open Source code is available at: https://github.com/CloudFerro/creodias-deployer |
| Partner | Pathway | Exploitation Path |
| UNITN | Open-Source Exploitation | We will update the services based on the feedback of the open-call users to ensure the projects can integrate the services effortlessly in their pipelines. After the AI4Copernicus project, we don't plan to update the services if not requested by a third-party user. If another user requires an update/modification, we are open to collaborating and cooperating with the interested entity, further improving the services or changing the architectures' scope. |
| SatCen | Open Source Exploitation<br>&<br>Institutional Exploitation | SatCen provided tools (e.g. bootstrapping services) developed in-house, adapted and fine-tuned for the AIODP integration. SatCen aims to explore the use of the outcome of the open call projects in its internal processing chains.<br>SatCen was mainly interested in testing its services' portability in different platforms (e.g., from a SatCen internal platform to the AIODP) and following the outcome of the open call projects (especially those relevant to security). AIODP platform would also support the standardisation of the preprocessing for future European projects.<br>The participation in AI4 Copernicus is interesting due to the outcomes of the open calls and the possibility of hosting added value services based on AI.<br>Satcen followed three projects: one on maritime, one in Climate security and one on humanitarian aid. The outcomes of these projects might be of interest. The SR4C3 (Super Resolution for Climate Crisis Context), a project led by SISTEMA, could open a close collaboration between the two organisations. |
| Equinor | Commercial Exploitation<br>Know How | Through AI4 Copernicus, Equinor has learned much about the ecosystem, and the open calls results. Results have been transformed internally to enhance their open innovation strategy and use platforms such as AI4 Copernicus and AIODP that can speed up the process.<br>Equinor has also identified start-ups that can or provide Equinor with services and expose the company to a wider audience. |
| BLUE-SIGHT (BLS) | Commercial Exploitation | In 3 years, BLS learned valuable information on the impact of EO and remote sensing can have on addressing key challenges such as climate change, particularly for cities. The plan is to expand BLS activities in the Smart and Climate Neutral Cities, using this knowledge, particularly regarding |

| | | |
|---|---|---|
| | | business models and financing. This will require keeping activities related to EO and AI and turning this knowledge into valuable advice for BLS primary markets: cities. |

# 3 Third-party exploitation strategies (thematic services)

D6.2 (use cases with predefined themes), D6.3 (small-scale experiments) and D6.4 (citizen-driven themes) provide an overview of the Open Call results. D7.3 focuses on the exploitation strategies developed by the OCs winners. However, the data available on these aspects are scarce, particularly if we compare them to technical data. Thus, the objective is to highlight critical elements for the AI services or applications powered by Earth Observation Data. Most of the insights were taken in the AI Ecosystem Forum (Athens 28-30 June 2023) in the open call session. We start by presenting the open-call winners (3.1). Then, we combine recent Venture Capitalist data, highlighting current business models in the EO industry with the ones developed by the OC winners (3.2). The main conclusions of the AI Innovation Index are then presented (3.3). A short conclusion capitalises on key learnings (3.4).

## 3.1 Third parties at a glance

AI4 Copernicus selected **29** projects, developing **26 assets** uploaded on the AIODP. The total should be higher as the 5th OC is not included in the 25 assets. Only 27 projects contributed to the development of assets. The two remaining ones were the citizen social challenges awarded by OC 2

The information in this section is extracted from the pitch given by **23** companies during the Athens events.

- Five are from the 1st open call (Sens4weed, Scaviho, Valens, SR4C3, Slide) developing use cases in 4 Industrial Domains: Energy, Security, Health, and Agriculture.
- Six are from the 3rd open call (ODFuse4Ship, EFSA, Planet, Optimal, Lobelia LIFE4ENVIRONMENT) focused on experiments with beneficiaries SMEs, startups, and spin-offs (single partner projects).
- Two from the 4th open call: Urbalitics, AI2Green f**or Use cases based on citizen-driven social challenges**, with beneficiaries consortia of high-tech & at least one low-tech SME, startups, spin-offs.
- Ten are from the 5th open call (Aqqa, Flora4cop, Lift Sentinel, Noemi, AIMPSI, Semilake, SandMap, THRUST, AI4west, AI-qa farm) **for Micro-Projects** addressed by **technology-advanced SMEs, startups, micro-enterprises from EU Member States and Associated Countries (single-beneficiary micro-projects)**

The following table shows the distribution of the OC within domains.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 101016798.

| Agriculture/forestry | | Energy | | Security | | Environment/Health | |
|---|---|---|---|---|---|---|---|
| Total | Projects | Total | Projects | Total | Projects | Total | Projects |
| 9 | SENS4WEEDS<br>SCAVIHO<br>OPTIMAL<br>ESFA<br>THRUST-4RESST<br>AI-QUA FARM<br>FLORA4Cop<br>PLANET<br>LIFE4ENVIRONMENT | 3 | Slide<br>ODFuse4Ship<br>AIMPSI | 4 | SR4C3<br>VALENS<br>AIEW<br>LIFT SENTINEL | 7 | LOBELIA<br>SANDMAP<br>URBALITICS<br>NOEMI<br>SEMILAKE<br>AI2 GREEN<br>AQQUA |

*Figure 5 Open calls winners pitching in Athens Ecosystem Event*

## 3.2 Business models current trends in the industry and open calls business models

According to Terrawatch's work on business modelling, we can distinguish two main categories of downstream business models.[8]

**The Decision Enablers** are companies creating tools to empower users of EO to build their applications, particularly in the climate change hot topic. They do so by providing making EO data accessible (through data marketplaces), fusable (through interactive platforms bundling EO, in situ and other data) and usable (through ready-made analytics). The Climate Solutions segment plays a significant role in bridging the gap between the availability of EO data and the creation of applications capable of providing insights. Earth Blox and Planet are two meaningful examples[9].

**The Insight Engines** are companies building applications powered by EO data, some of which are combined with aerial, drone, in-situ and other forms of data. Companies build applications delivering AI-driven services encompassing monitoring, on-demand, SAAS solutions and other business models. These companies focus on a specific vertical and are trying to develop a tailored solution to solve a particular problem instead of a multi-purpose application solving many issues.

Most of the open-call winners belong to the second category (23). The B2B model is prevalent in all verticals (23), followed by B2G (12). Two are also targeting B2C models. Several companies are trying to address different Business Models (B2B, B2G often).

[8] The State of Earth Observation for Climate: 2023 Edition Terrawatch newsletter

[9] https://www.earthblox.io/; https://www.planet.com/products/planetary-variables/

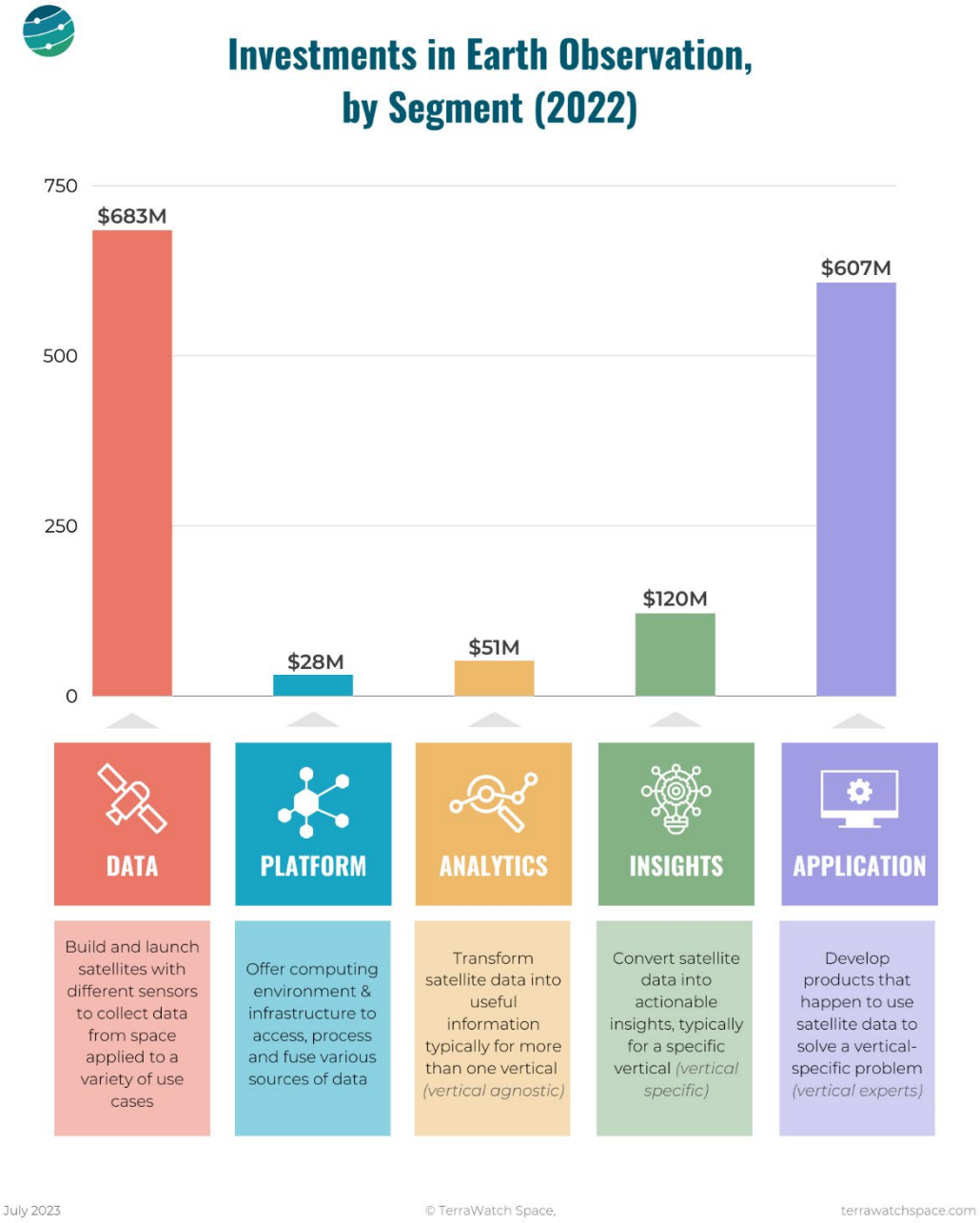



*Figure 6 Investment in Earth Observation by segment 2022 Terrawatch*

According to Terrawatch's recent studies, the insight, analytics and application companies seem to be receiving the lion's share of Venture capitalist funding.

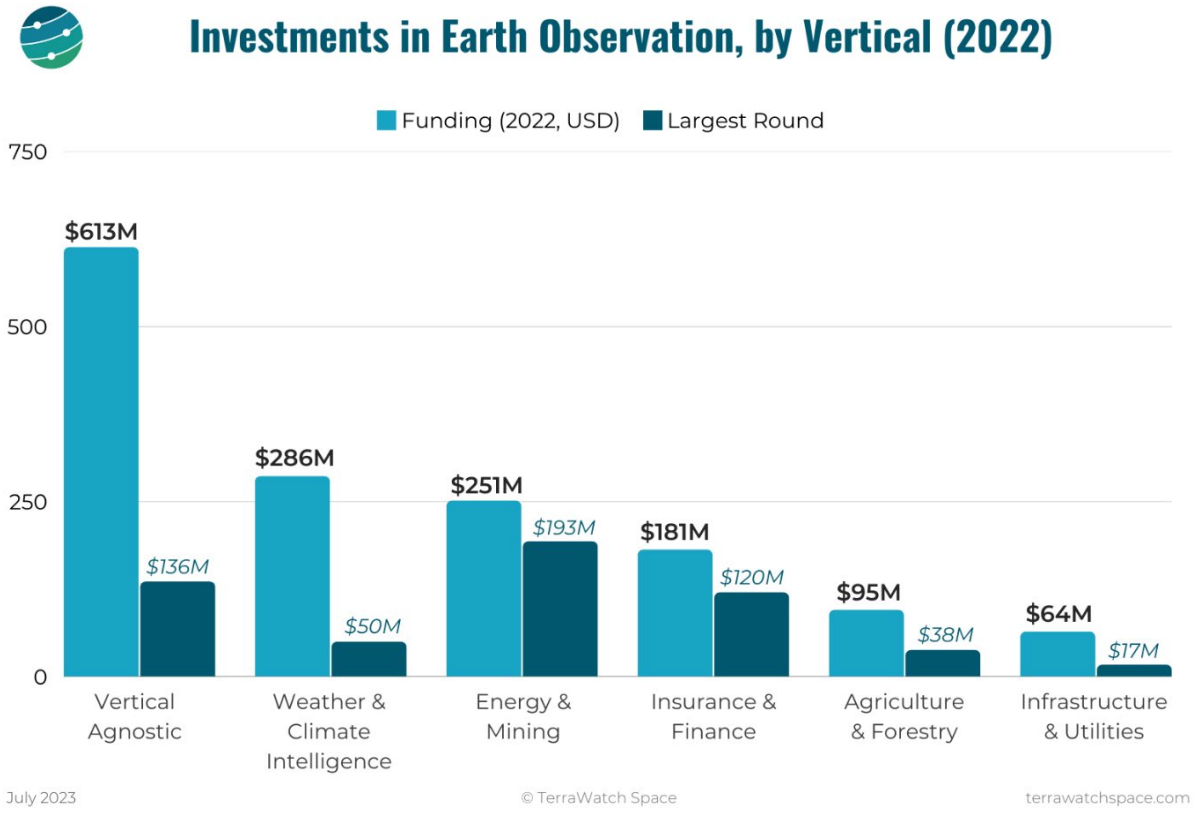



*Figure 7 Investment in Earth Observation by Vertical (2022) Terrawatch*

Over 40 per cent of investments were targeted at fundamentally horizontal companies, not verticalised like the rest of the investment.

### 3.2.1 AI4Copernicus open call agriculture and forestry business models

Problems to solve in this vertical encompass critical challenges such as.

- *How to enhance agriculture and forestry management, decrease intrants and improve yields and interventions in climate change and resource scarcity (see water management, for instance, of forestry management in a context of heatwave and dryness).*
- *How to develop biodiversity in the same context?*
- *How to improve aquaculture resources, location of fisheries and management?*

Business models are various and not completely set in stone as, frequently, different models are targeted (B2B, B2G). The most frequent ones are the following.

- We have only noted one “pure **Decision Enablers**” model based on a **Data As A Service to be incorporated in clients' crops models.**
- Business models consisting of **large-scale monitoring for crops or biodiversity** are developed. The first targets significant revenues through web-based interfaces and API to increase farmers' productivity. It looks pretty robust. The second is developing an ESG-driven approach, targeting companies needing ESG improvement scores for compliance, marketing, human resources and

product certification (biodiversity index). It's a buoyant market with growth forecasts triggered by regulation, particularly for companies and financial institutions.

- **On-demand solutions** are the third category. It looks like a consulting company supported by a robust technological backbone and AI tools. It's likely the less scalable as part of the service has to be tailored for each client and will not be automated.
- **Analytics services** are the last one.

### 3.2.2 AI4 Copernicus open call energy business models

Problems to solve are mostly to optimise energy efficiency, production and use.

- *Optimise shipping routes to save fuel and emissions by fusing multiple satellite data sources and using state-of-the-art methods in computer vision to obtain the best available operational predictions of ocean currents.*
- *Produce high-quality forecast images and higher ramp detection; nowcasting cloud covers to optimise solar farm production.*
- *Maximise adoption and benefits of Distributed Energy Resources while facing issues such as optimal locations, maintenance, and planning of solar installations*.

Business models are a combination of on-demand projects and analytics.

### 3.2.3 AI4Copernicus open call security business models

In the sample, 4 companies have worked on security issues. Problems to solve are the following.

- *Identify with great accuracy the impact of conflict on the environment by providing super-resolution algorithms.*
- *Identify non-compliant ships to allow a better intervention from SAR to identify and follow the vessels, decreasing the threat level regarding security and environment.*
- *Detecting forest areas, identifying illegal logging activities and monitoring forest conservation efforts*
- *Identifying water areas to detect dry regions or potential flood zones, assisting in disaster management, early information warning and water resource planning*.

In this domain, two business models fit with the Decision Enabler model, and the two other ones belong to the insight engine. Clients are B2G, and in the B2B sector, insurers and asset managers for the solutions developed to address climate and security risks.

### 3.2.4 AI4Copernicus health and environment business models

This vertical is the second one regarding project numbers, only overtaken by the agriculture domain. Problems to be solved are climate related.

- *Address air quality problems by providing easy access to Air Quality's numerous data and an urban monitoring and forecasting infrastructure scalable to any city.*
- *Addressing the heat urban island effect by an automated and augmented assessment of the situation combined with a selection in a portfolio of Nature-Based Solutions of the most efficient solutions to address the situation.*
- *Addressing the lake pollution by algae phenomenon by a semi-supervised representation learning-powered urban lakes and algae monitoring system that works with Sentinel 2 data and AI4Copernicus preprocessing service.*
- *Providing AI-powered technology for green space management to enhance their management and decrease the cost of energy and urban overheating*.

Most projects address urban challenges by developing AI climate applications powered by EO. Only one project seems to fit with the solution enabler models. Business models combine B2G and B2B, which is undoubtedly an excellent way to compensate for risks related to the lengthy public procurement process.

## 3.3 The AI/EO innovation index: key insights

### 3.3.1 An AI & EO Innovation Index

Aiming to empower further innovation in the AI and EO domain, we created an innovation index that will act as a reference for measuring the relative performance of the European AI and EO innovation ecosystem. The proposed AI & EO Innovation Index aims to support further the AI & EO Innovation Cycle in the course of the AI4Copernicus project and beyond by:

- Firstly, providing an overview of the ("perceived") AI & EO innovation performance across a selected sample of European SMEs that applied to the AI4Copernicus Open Calls: 3rd and 5th (N1=51 (3rd Open Call submissions in total), N2=67 (5th Open Call submissions in total).
- Secondly, assessing the relative strengths and weaknesses across European SMEs in AI and EO aims to help companies' national and regional stakeholders identify areas that need to be addressed to enhance AI & EO innovation in Europe.

**The methodological framework** for this innovation index utilised the European Innovation Scoreboard (EIS) approach. ESI assesses national innovation systems across countries (EU Member States, European countries and regional neighbours) and provides a comparative assessment of the Research and Innovation performance. Based on their scores, EU countries fall into four performance groups: Innovation leaders, Strong innovators, Moderate innovators and Emerging innovators.

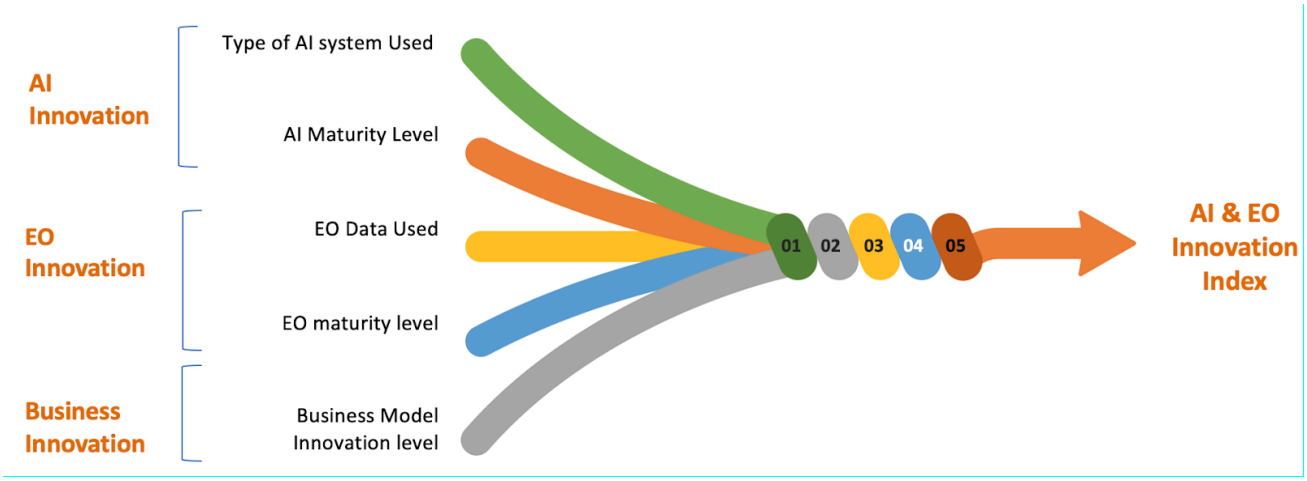



*Figure 8 The methodological approach of the AI & EO Innovation Index*

Adopting the EIS approach, we utilised three main types of criteria (AI Innovation, EO Innovation and Business Innovation) and five indicators for these criteria, as seen in the figure above. These indicators were included in the online submission platform, and companies self-assessed themselves.

| | Innovation Leaders (20% or more above the average) | Strong Innovators (above or close to the average) | Moderate Innovators (below average) | Modest Innovators (well below of EU avg) |
|---|---|---|---|---|
| **AI & EO Innovation Index** | 14% | **63%** | 24% | 0% |
| **AI Innovation Index** | 12% | **75%** | 14% | 0% |
| **EO Innovation Index** | 20% | **73%** | 8% | 0% |
| **Business (Model) Innovation Index** | 29% | **47%** | 0% | 24% |

*Table 9 The AI & EO Innovation Index for the 3rd Open Call Submissions (April 2022)*

**For the 3rd AI4Copernicus Open Call**, which involved single company (SME, start-ups, spin-offs that are "high technology" users) projects, conducting AI & EO experiments (with an end output at minimum TRL6), we received 51 submissions (April 2022) across Europe. As it can be seen in the figure above, when we examine the AI&EO Innovation Index, most **companies appear to be** "**strong innovators**", with 63% (above or close to the average), 24% being "moderate innovators" and only 14% are classified as "Innovation Leaders".

For the AI Innovation Index, we can observe **a similar pattern**, where the majority are perceived to be strong innovators (75%), followed by moderate innovators (14%) and innovation leaders (12%). For the EO Innovation Index we see that this pattern changes slightly as the majority consider themselves as strong innovators (73%), followed by innovation leaders (20%) and then moderate innovators (8% of respondents).

When we examine the Business Innovation Index that explores the business model innovation level of the applicants, we see a **different pattern**, as the majority perceive themselves as strong innovators (47%), followed by innovation leaders (29%) and modest innovators (24%).

Although innovators seem to be performing well in the AI-centric and data-centric innovation capacities at a high level, when the focus is on business aspects, a significant percentage appears to be lagging behind, expressing a potential area of improvement.

| | Innovation Leaders (20% or more above the average) | Strong Innovators (above or close to the average) | Moderate Innovators (below average) | Modest Innovators (well below of EU avg) |
|---|---|---|---|---|
| **AI & EO Innovation Index** | 4% | **72%** | 24% | 0% |
| **AI Innovation Index** | 12% | **57%** | 31% | 0% |
| **EO Innovation Index** | 9% | **46%** | 45% | 0% |
| **Business (Model) Innovation Index** | 25% | **43%** | 31% | 0% |

*Table 10 The AI & EO Innovation Index for the 5th Open Call Submissions (March 2023)*

For the **5th AI4Copernicus Open Call**, which involved single companies (SME, start-ups, spin-offs) being "high-technology" users and having a product in the market or conducting micro-projects (with an end output at minimum TRL-5), we received 67 submissions (March 2023) across Europe. For this open call, the AI & EO Innovation Index showed that **most companies are "strong innovators'** with 72%, 24% are "moderate innovators' ' and only 4% are classified as "Innovation Leaders".

We can observe a **similar pattern** for the **AI Innovation Index**, with lower rates. The majority are perceived to be strong innovators (57%), followed by moderate innovators (31%) and innovation leaders (12%). For **the EO Innovation Index, we see a different pattern** as the majority are strong innovators (46%), followed closely by moderate innovators (45% of respondents) and then innovation leaders (9%).

Regarding the Business Innovation Index, 43% perceive themselves as strong innovators, followed by moderate innovators (31%) and innovation leaders (25%). This indicates that although innovators seem to be performing well in the AI-centric innovation capacities at a high level when the focus is on data-centric and business aspects, a significant percentage appears to be struggling to improve as they are in the moderate innovator's position. This indicates **the need for strengthening data-centric and business aspects of the AI and EO domain for micro and small European innovators** to remain competitive.

### 3.3.2 AI & EO Innovation Index Findings

Our AI & EO Innovation Index findings for 2022 and 2023 AI4Copernicus open calls indicate that Europe needs **to strengthen this innovative industrial domain** to reap the innovation, economic, and societal benefits of the Earth Observation domain in the long run.

In particular, we can observe the following:

- Most European micro and small AI and Earth Observation companies tend to be **close to or above the average.**
- There is a tendency to **focus on the technological and data-centric** components rather than the business aspects (including their business model).
- New high-tech innovators focus on technology and data.

- More mature high-tech innovators adopt a more balanced approach around technology, data and business aspects.
- New innovators tend to be more confident than mature start-ups concerning their overall performance.

These findings indicate that the European AI and EO domain should provide support **across all phases** of the AI & EO innovation cycle, acknowledging the different needs that companies must go from discovery to research, to prototyping, and finally scaling up into a profitable new AI &EO innovation that the market adopts. Our findings and collaboration with AI & EO start-ups as part of the AI4Copernicus incubation process (WP6) indicated the need for diverse support **covering technical, business, and responsible AI & EO aspects**. Thus, monetary support should be combined with additional incubation and knowledge transfer support to cover the needs of European start-ups.

Finally, we believe it is essential to **benchmark** this innovative European ecosystem to understand its needs and evolution over time. We believe the proposed "**AI & EO Innovation Index" can create a benchmarking tool t**hat practitioners, researchers, policymakers, business leaders, and other stakeholders use to assess innovation in this highly evolving domain and measure progress in AI & EO innovation over time.

### 3.3.3 The Valleys of Death

The well-known "Valley of Death" (VoD) refers to the set of challenges that technology companies face during their early development phases. These challenges during the early stages of the start-up life cycle are usually attributed to **funding gaps, lack of support** (government support, incubation support, etc.), and **operational challenges**, among others.

From its conception, AI4Copernicus aimed to bridge this gap from "discovery to commercialization" for the AI and EO domain by providing both funding and innovation support on a technical, business and responsible innovation (AI4Copernicus incubation process) (Figure X).

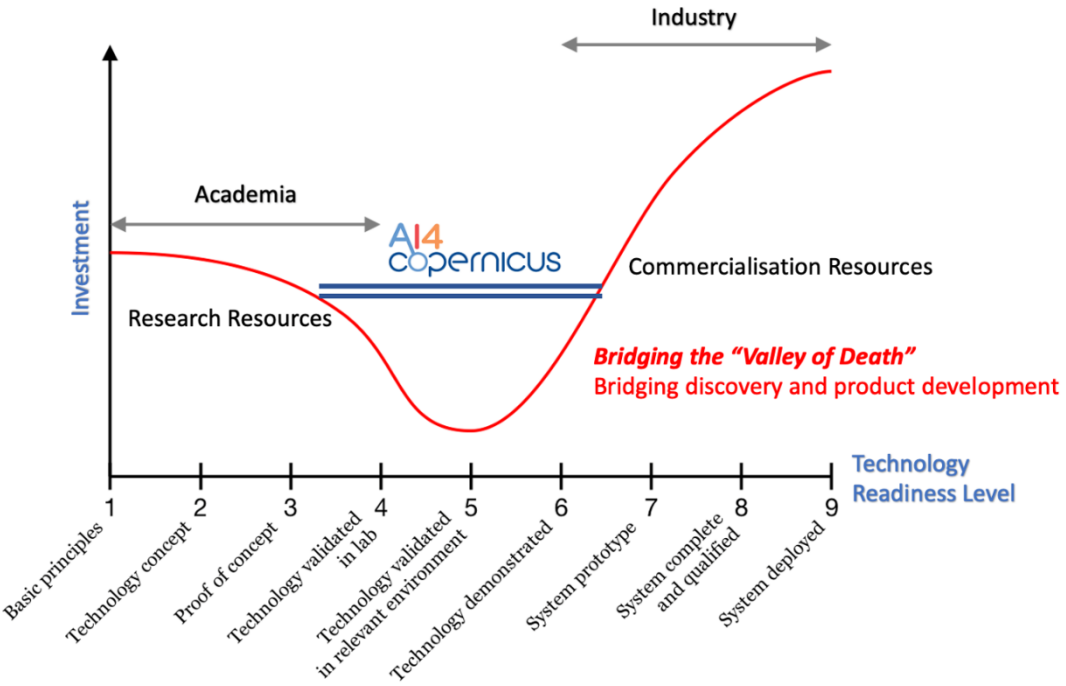



*Figure 11 Valley of Death: bridging discovery and product development*

However, during the project and after collaborating with numerous innovative EU start-ups in the AI & EO domain, in the course of WP6 and the AI4Copernicus Open Calls: 1st OC (6 projects – 12 start-ups). 3rd OC (8 projects – 8 start-ups), 4th OC (3 projects – 6 start-ups), 5th OC (10 projects/start-ups-SMEs) as well as during the Athens Event (Summer 2023) we identified the following aspects:

- **Market deployment risks** (and concerns) are evident (including preparation for a successful market deployment risks) for start-ups in this domain, including (1) achieving the right product-market fit (this appeared to be easier for the projects of the AI4Copernicus that involved a low-technology company to be part of the consortium, as this was usually the end-user or the means to reach end-users integrating market demands early in the project life-cycle) (2) adapting to market feedback from a business, technological and financial perspective, (3) designing and successfully executing their business plans.

- AI & EO is **a technology-push ecosystem**. The AI & EO domain is a technology-driven ecosystem with a significant **technology push rather than a demand-pull.** This creates a more substantial effort (marketing, sales, operational, etc.) for micro-companies, which creates a lot of uncertainty, especially for companies at TRL level 6 and beyond. This is a significant area of concern for all AI4Copernicus companies.

- **Tech-driven entrepreneurs in the AI & EO domain**: The entrepreneurs in this domain are mainly focused on the technological aspects of their innovations, and little emphasis is placed on the **business aspects of their innovations**. This common scenario creates risks and additional effort for start-ups to deploy and launch their products and services successfully.

- **Trustmarks and trust-assurance** mechanisms provide significant support for reaching a broader customer base.

- **Additional monetary support is needed** (a significant area of concern as stated by companies): the capital needs are increasing again during this phase, and adequate support must be provided so start-ups can go to the next stage of their business journey.

- **Additional non-monetary support** (a significant area of concern as identified by AI4Copernicus experts) is needed mainly business-oriented, including (business model, operational support, marketing, sales, legal, etc.)

Our findings denote the existence of **another "Valley of Death" (VoD 2),** which we called "Commercialisation Chasm" since companies need support to demonstrate commercial viability, performance, stability, and consistency to scale up (see Figure 12.) successfully.

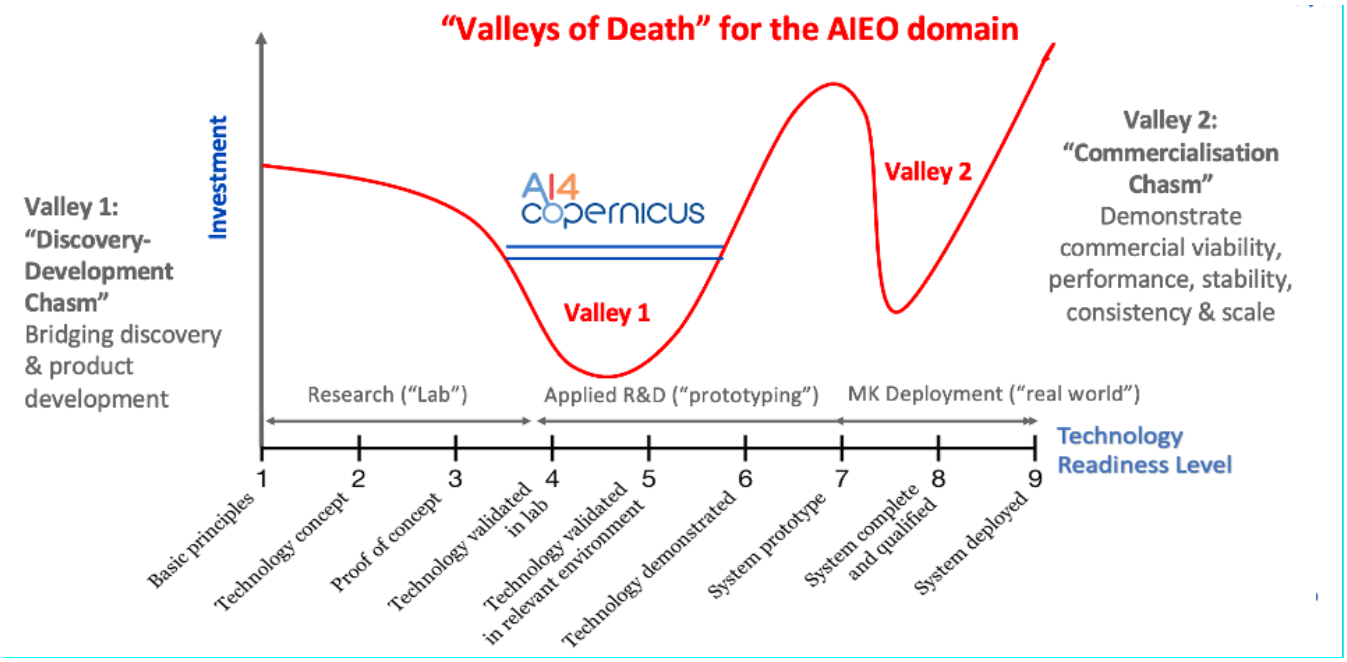



*Figure 12 the commercialisation chasm*

### 3.3.4 A VC Booklet for AI & EO

Aiming to support the AI4Copernicus start-ups across the next chasm in their life cycle, we designed and implemented a VC (Venture Capital) Booklet for the AI4Copernicus-funded companies and projects.

The booklet's design has been aligned with the VC expectations, enabling them to review, compare and focus on specific high-level areas of interest linked with each project's technological and business aspects. For this reason, we designed a novel template for an AI&EO "VC pitch" (see Figure X). The seven key areas of consideration were: (1). The company-specific overview, vision, mission, values and achievements; (2). Problem & Solution, (3) AI services & Innovative Aspects, (4) Target Audience, (5) Competitors, (6) Business Model & Business Targets (future targets), (7) Contact us information.


*Figure 13: AI4Copernicus VC Pitch template – example*

The information from the AI4Copernicus-funded projects was divided into five industrial domains: Agriculture, Security, Energy, Health, and Other Domains. A special section in the AI4Copernicus website was created, "Information for VCs", and the project VC pitches were included, as it can be seen below.


*Figure 14 AI4Copernicus website Information for VCs – Agriculture Example*

.

Additional information about the associated dissemination activities can be found in the WP7 Dissemination and Communication deliverable.

## 3.4 Conclusion Key learnings for third-party developments

At this stage, it is challenging to predict the sustainability of the projects after the European funding. The grant allocated is small; however, the effect can be significant for companies.
We can, nonetheless, highlight several conclusions.

### 3.4.1 A shift from a state-driven market to climate-related applications, in particular

Most of the project tries to address climate challenges, exemplifying the change of the EO market, usually driven by state applications. Even in the security domain, two projects address security risks triggered by climate change. Agricultural projects also clearly focus on increasing the efficiency and productivity of the farms in the same context.

### 3.4.2 Companies are close to commercial development but need support for business modelling and testing their value proposition.

Technology Readiness Level is not the proper tool to push for evaluating sustainability. Tools such as Commercial Inflexion Point (CIP) or the recent CARAT [10]framework developed by the US Department of Energy are better.
If we use the CIP tool, we can say that most of the projects are between CIP 6 (Solution is deployed at the commercial scale with at least two paying customers) and 7 (Solution is deployed at the commercial scale with a repeatable business model in a significant number of markets or with a substantial number of customers). CIP 7 seems, in most cases, the challenge that companies will address. But as the AI/EO Innovation index showed, support in terms of Business Models, Ethical aspects of the product and financing next steps remain.
Indeed, companies will deliver AI services powered by EO in various domains to European citizens. Some examples below

- Heat Urban Island (Urbalitics) is already working with two cities and has won a national contract with the Territory Agency of Italy).
- Lobelia Air presents a very similar dynamic in a different domain air quality.
- SISTEMA developed through the project SR4C3 a Super Resolution Algorithm for Climate Crisis Context.
- Sen4Weeds developed a solution for automated large-scale detection and mapping of weeds in agricultural fields, involving an end user, a Norwegian farm, as a test bed for the product.
- ESFA has recently been selected among Six start-ups by the business incubation center of ESFA.

### 3.4.3 However, making money for upstream applications is still a challenge.

Open call winners met in the AI Ecosystem Forum organised by AI4 Copernicus in Athens (more in section 4.2.3. One of the conclusions highlighted is the difficulty of making money from upstream applications for various reasons: the demand side is not yet sufficiently matured, and the customers' decision process could be lengthy and complicated. A scalable business through

[10] CARAT Commercial Adoption Readiness Assessment Tool

repeatable and scalable products is not accessible to develop as customers frequently need tailored applications that can require time to build, etc.
We will come back to this issue in section 4.2.3. However, one of the lessons learned through the AI4Copernicus call is to involve end users or customers in the process so as not to build only a technology solution but a buyable product. Using tools such as AI4Copernicus TrustMark, CARAT or the simpler CIP should also help to select companies with the potential to scale or help them scale their products.

Elemental Excelerator

**Commercial Inflection Points Scale**

Elemental Excelerator's Commercial Inflection Point (CIP) scale is a framework used to indicate a technology's degree of commercial readiness. Elemental's CIP is modified from the widely accepted Technology Readiness Levels (TRLs) and ARENA's Commercial Readiness Index (CRI) framework and is designed to categorize technology maturity as it fits within our evaluation criteria. The scale is numbered 1 to 8, with 8 being the most commercially mature.

| Commercial Inflection Point (CIP) | Description |
|---|---|
| 1 – Pre-Prototype / Research | Solution is in the early stages of ideation and development. Commercial applications have been identified and technology solutions are being explored. Founders are setting company values and starting to incorporate Diversity, Equity and Inclusion (DEI) principles. |
| 2 – Prototype Proof of Feasibility | Solution components and processes are being designed and tested in a lab setting. Company is confirming the technology viability of the prototype. |
| 3 – Prototype Beta Test | Solution prototype is being tested with a small potential customer base before being released to real world settings. Company is ready to deploy solutions for demonstration in an operational environment, but commercial and operational drivers have yet to be tested in the field. Implications and potential benefits for frontline communities begin to be assessed. |
| 4 – Initial Pilot | Solution is deployed in small-scale pilots in a real-world setting. Company is using early technical and commercial results to drive interest for first commercial deployment and inform further product development and market intelligence. Company may not be paid from pilots. Company is listening to communities to better understand the unique history and context of the places they want to deploy. |
| 5 – Initial Commercial-Scale Deployment | Solution is deployed in a commercial-scale project with a customer - and community partner - for the first time. Company may not be paid from deployments. Company is using results to prove viability in real world settings and generating additional commercial interest. |
| 6 – Two+ Commercial-Scale Deployments | Solution is deployed at the commercial scale with at least two paying customers. Company is using technical and financial results to further inform growth plans in terms of team/board composition incorporating DEI principles, product, market, and sales strategy. |
| 7 – Deployed at Scale | Solution is deployed at commercial scale with a repeatable business model in a significant number of markets or with a significant number of customers. The business is continuing to expand and grow revenue. |
| 8 – Market Leader | Solution is delivering significant value to customers. The sales process is repeatable with predictable revenue in multiple customer segments and/or multiple markets. The solution is gaining market share and the company is broadening its product offerings and customer support services. An engine to generate new jobs and measurable climate benefits has been created. |

*Figure 15 Commercial Inflection Point Framework*

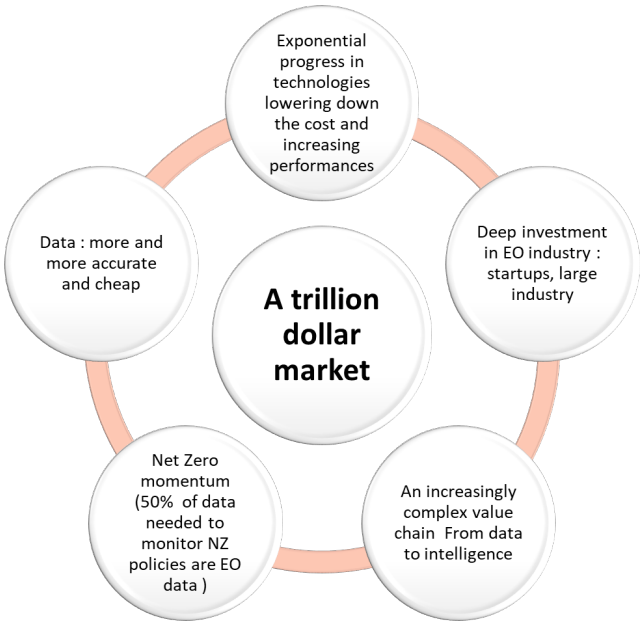
# 4 AI4Copernicus Sustainability Plan Towards an AI/EO node on the AIODP

This section **proposes several scenarios for building an AI/EO node** for the next phase of the AIODP. In a context of fierce global competition, the consortium firmly believes in the added value of this node to significantly **contribute to a more EO user-centric approach.**
We start by answering the **Why** of an AI/EO Node; what are the challenges and the needs (Section 4.1.) where we review them. We then work on the **What** by proceeding to an extensive benchmark and presenting different standpoints (from the partners and the OC winners). We present in section 4.3 **what pathways** could be taken for developing the AI/EO node and can be further developed by the follow-up project, Deploy AI.

## 4.1 Why: challenges and needs

### 4.1.1 Catching up the iPhone moment for EO

D7.2 pointed out the "iPhone moment "for the Earth Observation market due to several trends summarised by the diagram below and the quote of Peter Platzer




"The disruption from mainframe computers to personal computers and, eventually, to the internet is an almost perfect analog to what is happening in the space industry. We have a few dominant internet players today, but there are almost no companies that don't use the internet and computers. ***I think that is where the space industry is headed: there will be some large players, but the use of space will be widely distributed because access to it is becoming more regular."***

*SPIRE CEO Peter Platzer*

*Figure 16 Key trends for the EO market. Blue-Sight*

However, "EO remains a niche market (4% of the space market), and **its growth** (7% forecast) **is driven by downstream data and services**, as upstream and other activities will deliver limited growth. While the global data market is highly consolidated, the service market is more fragmented and in consolidation[i11]."
Catching up with the iPhone moment and mainstreaming EO must address at least three main blocking points.

[11] ESA Week 2021 October 11-15, 2021, Summary and recommendations to ESA

- **Address the current fragmented landscape and turn upside down the dominant supply side and techno-driven reasoning with a user-centric approach**, trying to solve concrete problems.
- **Addressing disruptive innovation**, which will be soon on the market, such as EO large language or foundational models.
- Developing the **research agenda and the education /training** as **competencies are currently in short supply.**

### 4.1.2 Catching up with a user-centric approach for EO and AI

The ESA program for mainstreaming EO and AI is powerful and will bring excellent results. We can highlight several programs, such as:

- **AI4EO Solution Factory** (https://www.ai4eo-factory.de/en-index.html ) is a collaborative environment for developing customised solutions using Earth Observation (EO) data and Artificial Intelligence (AI) tools by employing existing modules of the AI4EO Solution Factory. Potential customers are active in one of the many application areas of EO data, such as agriculture, forestry, fishery, natural disaster forecasting/ monitoring and urban planning. They have a problem statement and seek solutions based on EO data and AI tools. The team will develop a business case around this problem statement and develop a solution jointly with employees from the customer.
- **Destination Earth (EC) and Digial Twin Earth (ESA)** and its precursors (Antarctica, Food systems, Hydrology), etc.
- **ESA Virtual Laboratory and open innovation tools** platform services are primarily intended for scientists to share data resources and create enhanced research environment.
- **The ESA Earth Observation Training Data Lab (EO-TDL[12])** will address the challenges raised by the scarcity of suitable and accessible training datasets. EO-TDL will provide a cloud repository to create, share, and improve training datasets and ML/DL algorithms. The goals of EO-TDL are:
  - Host, import and maintain a wide range of dataset types: training, validation, test, benchmark, and reference datasets (in-situ data, product validation datasets).
  - Offer integrated open-source tools compatible with the major ML/DL frameworks to develop and export processing pipelines for Extract Transform Load (ETL) operations, data ingestion, model training and inference.
  - Enable the description, versioning and tracking of data using the Spatio Temporal Asset Catalog (STAC) to guarantee data discoverability and accountability.
  - Allow data exploration to uncover biases, detect anomalies, and verify assumptions, maximising the understanding of the data (Exploratory Data Analysis-EDA).
  - Build a centralised Feature Store to access, search, and create EO data-derived features and serve them at training and inference time, thus increasing model efficiency.
  - Enable automated data quality mechanisms through deterministic and non-deterministic testing.
  - Deploy a containerised multi-GPU environment for distributed training processing.
  - Provide interoperability with third-party platforms, such as Radiant Earth MLHub.
  - Implement accessibility at multiple levels using user interfaces, web APIs, CLIs and Python libraries.

[12] EARTH OBSERVATION TRAINING DATA LAB Draft Brochure 22 November 2022

- **Scout MISSIONS**, a new family of small satellites, delivers value-added science by miniaturising existing space technologies or demonstrating new observing techniques. The overarching aim is to achieve this with an extremely agile and low-cost development process while offering the space industry and academia a proactive role. Three key aspects define a Scout mission:
  - It must deliver innovative science and technology.
  - It must cost less than €30 million.
  - It must be developed within three years, from kick-off to launch.
  - **The New Copernicus Data Ecosystem and Digital Twin Earth are two well-known flagships.**

These tools and programs will deliver impressive results and facilitate EO mainstreaming. However, **recent papers highlight users' challenges** with a still scattered ecosystem and an exponential growth in data due to the launch of multiple satellites and the imagery progress.

*Earth observation (EO) is a prime instrument for monitoring land and ocean processes, studying the dynamics at work, and taking the pulse of our Planet. A large variety of sensor data (active /passive / of many resolutions) are nowadays accessible to researchers, agencies, and the general public. However, a final barrier remains the need for technology to convert the* ***enormous quantities of raw EO data generated daily into the valuable information necessary for making decisions and taking concrete action****. The ability to extract meaningful information from raw EO data is an essential prerequisite to achieving significant impact, be it in monitoring or documenting (e.g. progress towards the United Nations's sustainable development goals), predicting and issuing timely warnings.*(e.*g. related to future natural disasters [and need for emergency evacuations), or projecting the effects of human actions and natural processes on nature and society.*[13]

*Today's EO exploitation platforms are limited to the processing functionalities and datasets they provide, and there is no single platform that provides all datasets of interest. Thus,* ***enabling cross-platform (federated) analytics is crucial to make EO technology easily accessible to everyone****. We envision AgoraEO as an EO ecosystem for sharing, finding, composing, and executing EO assets, such as datasets, algorithms, and tools. Making AgoraEO a reality is challenging for several reasons, mainly because the ecosystem must provide interactive response times and operate seamlessly over multiple exploitation platforms****. An open, unified EO ecosystem*** *would foster innovation and boost EO data literacy for the entire population.*[14]

**The challenge to mainstream EO is to create this federation of platforms through APIs and user interfaces**. **AI4 Copernicus** started to build components of such federation with, on the one hand, the **integration between the AIODP and CREODIAS**, now the Copernicus Data Ecosystem and on the other hand, **the development of the Earth QnA Engine**.

Other projects contribute significantly to this challenge, like *the openEO API (Application Programming Interface) standardising EO-related contracts between local clients (R, Python, and JavaScript) and cloud service providers regarding data access and processing, simplifying their direct comparability. Independent of the providers' data storage system, the API mimics the functionalities*

[13] Artificial intelligence to advance Earth observation: a perspective arXiv:2202.07242v1 [cs.CV] 15 Feb 2022
[14] AGORA-EO: A UNIFIED ECOSYSTEM FOR EARTH OBSERVATION– A VISION FOR BOOSTING EO DATA LITERACY Berlin Technical University 2021 Big Data for Space Conference

*of a virtual EO raster data cube.* [15]*;*. **This user-centric approach is crucial to developing large open science projects and mainstream EO usage, enabling Europe to benefit from disruptive innovation**.

### 4.1.3 Catching up with disruptive innovations and research programs

#### *4.1.3.1 Toward large language and foundational models for EO*

Over the last years (and months for AI), groundbreaking progress has been made in two domains: language and foundational models for AI, imagery, and computer vision for EO. The two areas will likely converge to create specific large language or foundational models.

Through ESA and EC, Europe is implementing ambitious projects such as Destination Earth that will bring significant impacts.

However, as shown in the next page diagrams, the competitive landscape does not favour Europe. The diagram shows the incredible pace of the adoption curve for Chat GPT4 and the multiplication of specialised LLM, which makes the creation of foundational models even more important to set up a user centric EO Ecosystem.

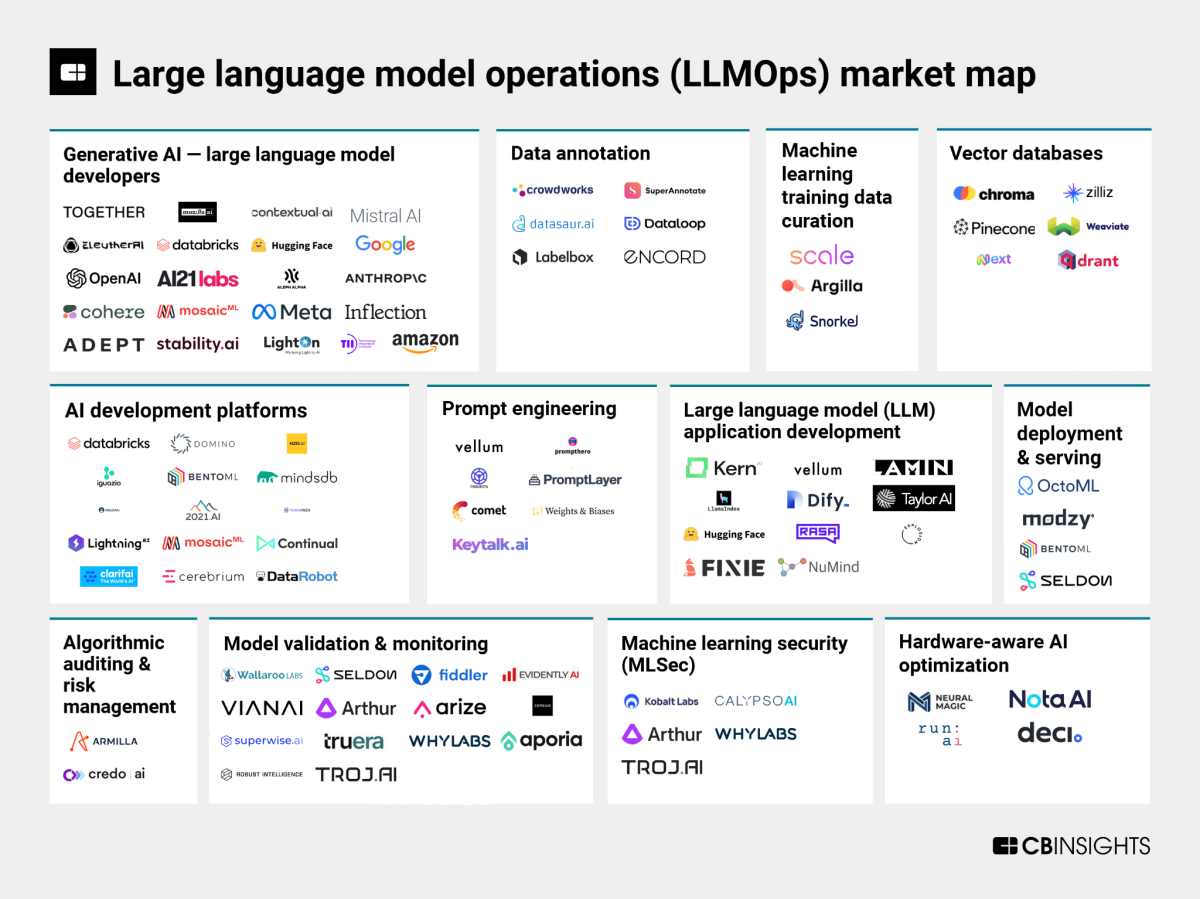



Figure 17 Large Langage Model Operation market map. CB Insights Report 2023

[15] The openEO API–Harmonising the Use of Earth Observation Cloud Services Using Virtual Data Cube Functionalities Remote Sens. 2021, 13, 1125. https://doi.org/10.3390/rs13061125

## Generative AI — large language model developers

The generative AI — large language model developers market offers foundation models and APIs that enable enterprises to build natural language processing applications such as content creation, summarisation, classification, chatbots, sentiment analysis, and more. Enterprises can fine-tune and customise these large-scale language models — pre-trained.[16]
Those companies are developing closed (i.e. Open AI) or open models (Huggin Faces or Stability AI). They receive the lion's share of VC funding, as shown in the diagrams next page.

## AI development platforms

The AI development platforms market offers solutions that serve as one-stop shops for enterprises that want to develop and launch in-house AI projects. Vendors in this space enable organisations to manage all aspects of the AI lifecycle — from data preparation, training, and validation to model deployment and continuous monitoring — through a single platform to facilitate end-to-end model development. Some vendors offer "drag-and-drop" interfaces or "plug-and-play" solutions that enable teams without in-depth AI expertise to build AI projects.

These two segments received most of the equity funding (85%).

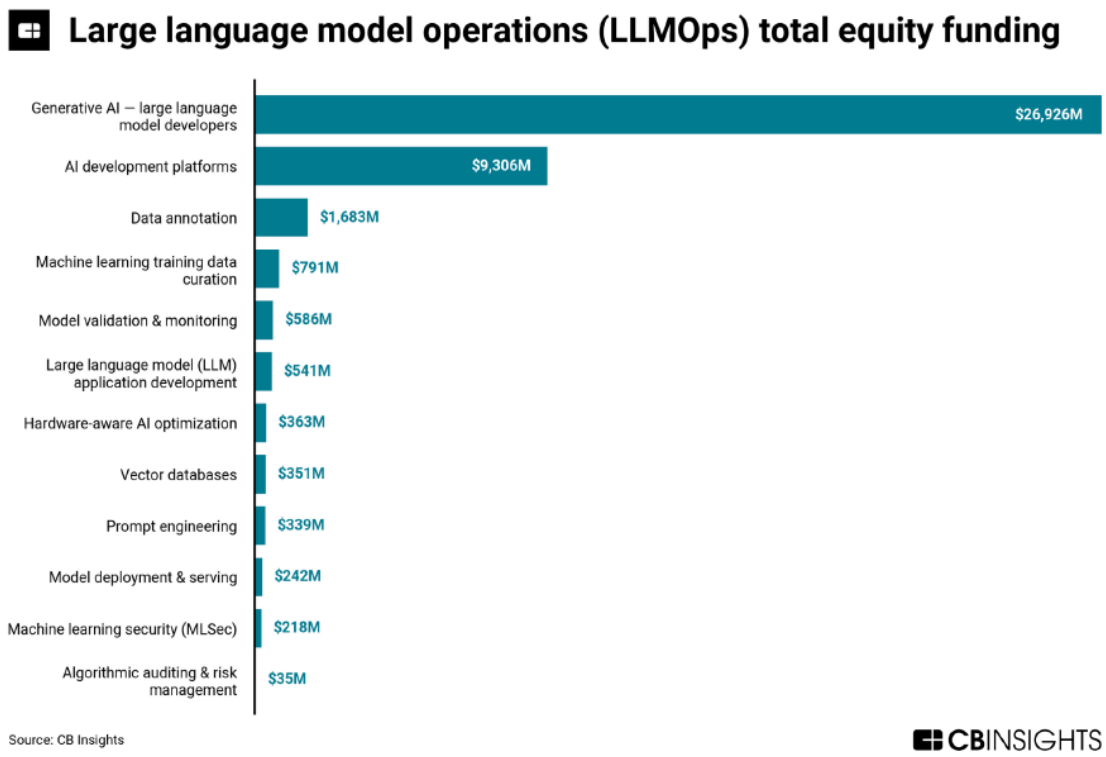



*Figure 18 Large language Model Operations total equity funding. CB Insights*

CB Insight is an American Company. Data can have a US bias oriented. However, when it comes to generative AI unicorns, the bias is limited as we can find the names of the big players (Open AI, Huggins Face, Stability AI, Inflection, etc.). Only one company is not located in the USA but in Israel.

[16] The large language model operations (LLMOps) market map CB Insights report2023

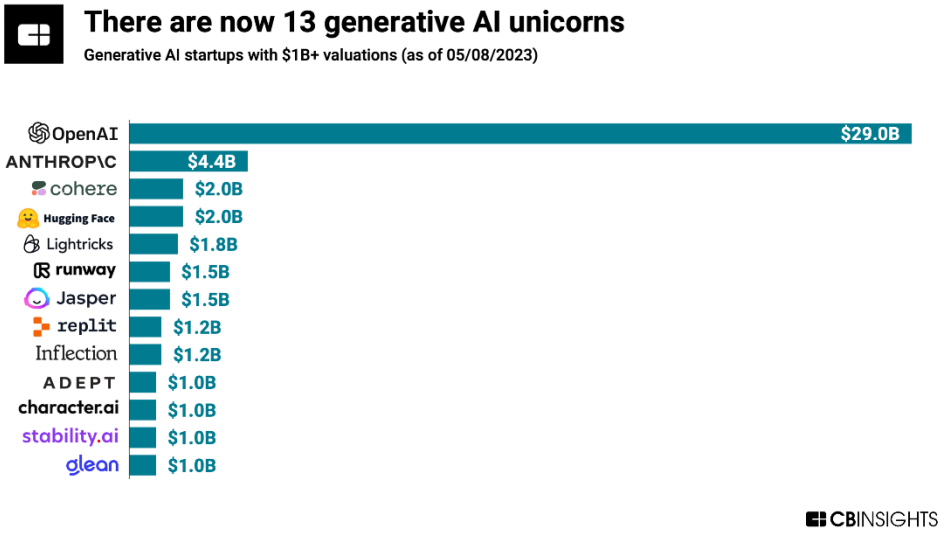



*Figure 19 Generative AI Unicorns CB Insights*

However, few of these companies seem to be targeting EO foundational models.
AI, combined with EO data, is powerful in addressing critical challenges (climate change, security etc.). Large-scale initiatives would likely flourish to develop the next big thing in EO or related applications.

### *4.1.3.2 Two major initiatives from the US*

**Nividia New supercomputer will create a Digital Twin Earth 2**

According to Nvidia, we need more data, better modelling, and faster computers to know where and how to take action on climate change. That's why the company is building what it calls "the world's most powerful AI supercomputer dedicated to predicting climate change." The system will be called Earth-2 and created using Nvidia's Omniverse, a multi-GPU development platform for 3D simulation based on Pixar's Universal Scene Description. In a blog post announcing Earth-2 late last week, Nvidia's founder and CEO Jensen Huang described his vision for the system as a "digital twin" of Earth.

Plenty of climate models already exist. They quantify factors like air pressure, wind magnitude, and temperature and plug them into equations to get a view of climate patterns in a given region, representing those regions as 3D grids. The smaller the region, the more accurate a model can be before becoming unwieldy (in other words, models must solve more equations to achieve higher resolution, but trying to take on too many equations will make a model so slow that it stops being useful). This means most existing climate models lack both granularity and accuracy. The solution? A bigger, better, faster computer. "Greater resolution is needed to model changes in the global water cycle," Huang wrote. "Meter-scale resolution is needed to simulate clouds that reflect sunlight back to space. Scientists estimate that these resolutions will demand millions to billions of times more computing power than what's currently available."

Earth-2 will employ three technologies to achieve ultra-high-resolution climate modelling: GPU-accelerated computing; deep learning and breakthroughs in physics-informed neural networks, and AI supercomputers—and a ton of data. The ultimate aim of this digital twin of our Planet is to spur action that will drive meaningful change, both in terms of mitigating the negative impacts of climate change on populations and mitigating climate change itself. Extreme weather events like hurricanes, wildfires, heat waves, and flash floods are increasingly taking lives, damaging property, and forcing people to flee their homes. If we could accurately predict these events much further in advance, those headlines would change.

Nvidia has not shared a timeline for Earth-2's development nor when the supercomputer will be ready to launch. But if its Cambridge-1 supercomputer for healthcare research is any indication, it **won't be all that long; Cambridge-1 took just 20 weeks to build and is ranked as one of the 50 fastest computers in the world**.[17]

### NASA/IBM foundational model[18]

"NASA and IBM have created an AI Foundation Model for Earth Observations, using large-scale satellite and remote sensing data, including the Harmonized Landsat and Sentinel-2 (HLS) data. Both organisations actively contribute to the global mission of promoting knowledge sharing and accelerating innovations in addressing critical environmental challenges by embracing open AI and science principles. With Hugging Face's platform, they simplify geospatial model training and deployment, making it accessible for open science users, startups, and enterprises on multi-cloud AI platforms like watsonx. Additionally, Hugging Face enables easy sharing of the pipelines of the model family, called Prithvi, within the community, fostering global collaboration and engagement.

The goal of the NASA/IBM work is to provide an **easier way for researchers to analyse and draw insights from large NASA datasets related to Earth processes.**

"We believe **that foundation models have the potential to change the way observational data are analysed and help us to better understand our planet**," says NASA Chief Science Data Officer Kevin Murphy. "**And by open sourcing such models and making them available to the world, we hope to multiply their impact**."

Foundation models have the potential to play a pivotal role in understanding our Planet's interconnected processes and the climate effects of ongoing natural and human-caused changes. Foundation Models pretrained on Earth observation data can accelerate the analysis of tremendous amounts of data in two primary ways.

First, FMs do not need large training datasets, which can be laborious and resource-intensive to create. Training FMs on much smaller datasets can save time and money. Second, FMs can reduce

[17] https://singularityhub.com/2021/11/17/nvidias-new-supercomputer-will-create-a-digital-twin-of-earth-to-fight-climate-change/
[18] https://www.earthdata.nasa.gov/news/impact-ibm-hls-foundation-model?utm_source=substack&utm_medium=email

redundant efforts to build downstream applications, which use FM output to perform a specific task, such as tracking changes in land use or monitoring natural disasters".

These two examples show the dynamic attached to EO and AI and the focus on creating open tools that will allow researchers and developers to fine-tune Foundational Models and fully exploit the full spectrum of EO Data.

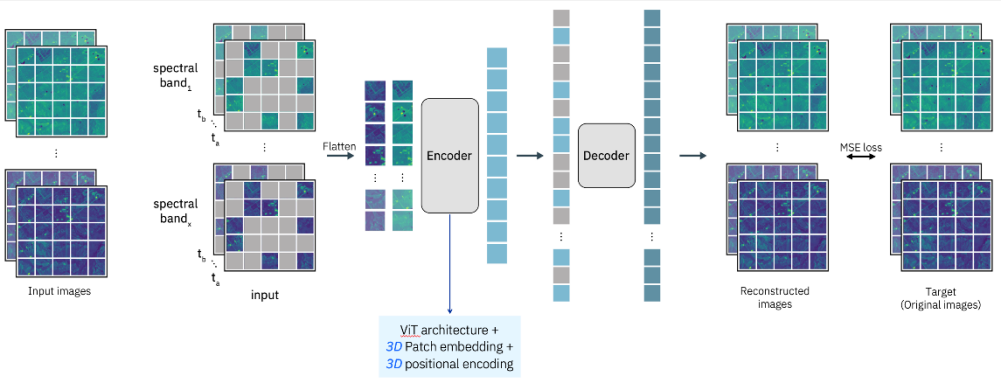
Model card Files and versions

https://huggingface.co/ibm-nasa-geospatial

ibm-nasa-geospatial Prithvi-100M
Pytorch Geospatial Temporal ViT Vit
License: apache-2.0

Model and Inputs

Prithvi is a first-of-its-kind temporal Vision transformer pre-trained by the IBM and NASA team on contiguous US Harmonised Landsat Sentinel 2 (HLS) data. The model adopts a self-supervised encoder developed with a ViT architecture and Masked AutoEncoder (MAE) learning strategy with an MSE loss function. The model includes spatial attention across multiple patches and also temporal attention for each patch.




The model accepts remote sensing data in a video format (B, C, T, H, W). Note that the temporal dimension (T) is essential in this application and not present in most other works around remote sensing modelling. The ability to handle a time series of remote sensing images can benefit various downstream tasks (e.g. Burn Scars segmentation, Flood Segmentation, Land Cover Classification). The model can also handle static imagery, which can be fed into the model with T=1.

**Pre-training**

The model was pre-trained with NASA's HLS V2 L30 product (30m granularity) from the contiguous United States.

### 4.1.4 Towards a collective research agenda

The Towards a Collective Agenda on AI for Earth Science Data Analysis paper[19] develops six main axes of research that can contribute to making the EO more user-centric. Three of them are at least well aligned with the AI4 Copernicus project.
We highlight from the paper these research themes.

#### *4.1.4.1 REASONING AND HUMAN-MACHINE DIALOGUES*

"*Current research at the interface between machine learning and remote sensing largely focuses on directly recognising materials and objects or estimating geophysical parameters. The reasoning goes beyond the recognition concept and aims at mimicking how people think and learn. It is centred around tasks such as induction, deduction, spatial and temporal reasoning, and structural inference. The work mentioned above showcases how spatial relational reasoning helps improve the semantic understanding of remote sensing images, and many other problems may also benefit from visual reasoning. One exciting example is temporal reasoning for analysing multitemporal data/aerial videos, e.g. for event recognition. This is a new, exciting field where one is concerned with understanding complex events being imaged or filmed, such as cultural events, manifestations, or locating people in distress. Using reasoning enables understanding if a person on a roof during a flood needs help or if a video of a crowd is related to a pacific or violent manifestation*."

#### *4.1.4.2 EXTREMELY MULTI-MODAL REMOTE SENSING*

*Remote sensing is not restricted anymore to observation with airborne or satellite sensors. Nowadays, we can monitor our Planet's health and status with social media data, socio-economic indicators, all kinds of imagery, audio, text, and satellite imagery*. This direction highlights the challenge of fusing data from various sources, formats, etc., in a user-centric approach. That was one of the challenges some AI4 Copernicus Open Call winners faced.

#### *4.1.4.3 INTERPRETABLE AND EXPLAINABLE MACHINE LEARNING*

*Using machine learning for scientific applications aims at acquiring new scientific knowledge from observational data. Additionally, to the accuracy of the results, their scientific consistency, reliability, and explainability are of central importance. A prerequisite to achieve those is to design models that can be challenged; in other words, to create models whose inner functioning can be visualised, queried or interpreted*.

**The challenge is high as it will give trust to the market in the solutions developed. AI4 Copernicus developed a pioneering approach related to ethics**.

## 4.2 What could be an AI/EO node?

The concept of an AI/EO Node comes **from a strategic discussion for the future of the AIODP**. Is the AIODP a general platform without specialization? Or does the AIODP, to thrive in a very congested platform market, must focus on specific sectors, as the environment from one market to

[19] Towards a Collective Agenda on AI for Earth Science Data Analysis Devis Tuia, Senior Member, IEEE, Ribana Roscher, Member, IEEE, Jan Dirk Wegner, Nathan Jacobs, Senior Member, IEEE, Xiao Xiang Zhu, Senior Member, IEEE, Gustau Camps-Valls, Fellow, IEEE. arXiv:2104.05107v1 [ cs.CV] 11 Apr 2021

another is very different for deploying applications.? And then creates some verticals where the AIODP can reach a critical mass.

The AI4Europe project provided one exciting answer**: the AI Nodes**, which can balance the two strategies. Thus, one of the pathways for sustainability for AI4 Copernicus was to study the strategic positioning and the features of an AI/EO Node.

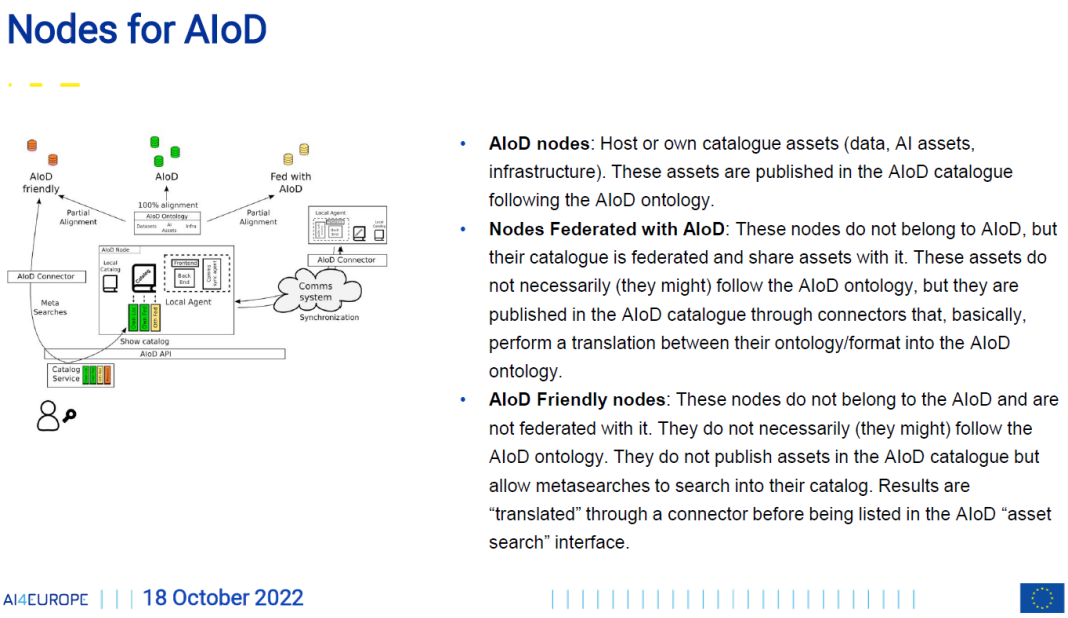



*Figure 20 Nodes for AIOD AI4 Europe presentation*

AI4 Copernicus Exploitation Working Group conducted a benchmark analysis to identify the main projects and assess the positioning and added value of an AI/EO node (section 4.2.1). We then interviewed all the partners to prioritise the most essential features an AI/EO node should cover (section 4.2.2). These analyses were then shared with the open calls winners during the Athens Ecosystem Forum, where three world cafes were organised (Section 4.2.3). We summarise in Section 4.3 the main conclusions.

### 4.2.1 Benchmarking

#### *4.2.1.1 Benchmarked projects*

We have worked on the following projects.

- **The news Copernicus Data ecosystem** https://dataspace.copernicus.eu/

- **Europe Data Cube** https://eurodatacube.com/. Europe Data Cube is calling itself the Earth Observation Information Factory. They propose a set of services.
    - Global Archives of Analysis Ready Data
        - Open satellite missions - Sentinel, Landsat, MODIS, etc.
        - Commercial VHR datasets - PlanetScope, Pléiades, SPOT, etc.
        - Earth System Data Cube

        - ESA and Copernicus Climate Change Initiative
        - User contributed content.
        - Data fusion to combine various datasets
    - Access and Analysis
        - Sentinel Hub - Cloud API to most important EO datasets for ad-hoc access, interactive exploration, and integration in 3rd party applications
        - Batch Processing - heavy-duty processing tasks for large-scale analysis and machine learning
        - xcube -customisable analysis and processing based on xarray technology
    - Storage and Distribution
        - geoDB for geospatial vector data and Sentinel Hub for raster data
        - Control the distribution to specific users or user groups.
    - Exploitation
        - EOxHub Workspace – managed compute and storage environment to run Jupyter Notebooks and to host own applications.
    - Collaboration
        - Marketplace for free or revenue-generating options to share data, applications and algorithms.

- **Git hub satellite-image-deep-learning** https://github.com/satellite-image-deep-learning?utm_source=substack&utm_medium=email. The satellite-image-deep-learning organisation provides resources on deep learning applied to satellite and aerial imagery and hosts the following repositories:
    - annotation: includes information on the annotation of datasets
    - datasets: lists many datasets
    - model-training-and-deployment: lists information on the training and deployment of deep learning models
    - software: for working with satellite & aerial imagery data & datasets
    - techniques: for deep learning with satellite & aerial imagery
- Destination Earth https://destination-earth.eu
- Earth 2 https://www.nvidia.com/en-us/high-performance-computing/earth-2/
- NASA/IBM Foundational models
- The AIODP (https://www.ai4europe.eu/); We have worked with the version online until October 2023. The announced new platform, developing new features and merging the catalogues (the one from the AIODP website https://www.ai4europe.eu/research/ai-catalog and the AI4Experiment catalogue https://aiexp.ai4europe.eu/#/home) wasn't available at the time of the analysis.

The benchmark was mostly conducted through the Internet. Some projects are running (e.g. Europe Data Cube), others are starting (e.g. Copernicus data ecosystem, NASA) or restructuring their services (e.g. AIODP). Earth-2 seems not to have started, probably because the announced supercomputer has not yet been delivered.

#### *4.2.1.2 Benchmarked features*

We selected several features because they are present in some platforms or represent expectations from a user standpoint. A questionnaire filled out by the AI4 Copernicus exploitation group members (11 partners) was conducted, followed by deep interviews (2 hours) to select the most essential features.

Some are technical and concern the access to EO data and their processing (access to EO data (access to EO data set via Diases and their services, storage and computing, repository, experiment, and search). The others are more related to non-technical aspects, such as research and innovation projects (existence of open sciences or innovation challenges), ethical aspects, matchmaking (competencies), marketplace (how to sell or buy assets), education or technology transfer. We have tried to assess whether the feature exists (Yes for existing, no for not and? if it is difficult to know) in the benchmarked projects and platforms.

The results are presented in the next page.

#### 4.2.1.3 Results

| Features/services | DAS ecosystem | Europe Data Cube | AI4Europe | AI4 Copernicus | Github satellite-image-deep-learning | Destination Earth | Earth2 | NASA/IBM foundational model |
|---|---|---|---|---|---|---|---|---|
| Access to EO data set via Diases services | Yes | Yes | No | Yes | Yes, through a list | Yes, Destination Earth Data Lake | Yes Nvidia Omniverse | Yes |
| Storage and computing | Yes | Yes | No | Yes | No | Yes | Yes | No |
| AI/EO repository (tools/algorithms) | No | Yes market place | Yes, but general (not related to EO) | No | Yes (tools) | Yes, basic software suite service for local data exploitation, data and software catalogue services, 2D/3D data visualisation service. | No | No |
| Testing/Experimenting AI4 Experiment | No | No | Yes | No | No | No | No | No |
| Research and innovation challenges/projects | No | No | Yes small scale | Yes | No | ? | No | Yes, open science + Huggin Face |
| Ethical Trustworthy | No | No | Yes but general | Yes | No | ? | ? | ? |
| Matchmaking | No | No | No | No | No | No | No | No |
| Search | ? | ? | ? | Yes | No | ? | ? | ? |
| Market place | Yes | Yes | No | No | No | No | ? | ? |
| Education | No | No | Yes | Yes | No | ? | ? | ? |
| Technology transfer | No | No | No | Partially | No | No | No | No |

*Figure 21 Benchmark European and international projects*

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 101016798.

The main findings coming from this benchmark are the following:

- There is a **large diversity of models and platforms** between those providing data, processing services (Copernicus data Ecosystem), repository or marketplace (Git hub, European Data cube), or large-scale projects developing an entire ecosystem (Digital Earth, Earth-2, NASA/IBM/Huggin Faces and more data-driven Copernicus Data Ecosystem).

- Prominent players, mostly public in Europe (EC, ESA) and more private overseas, are undertaking **large-scale projects with massive investments to develop on top of EO data algorithms and models**. Most intend to address climate change through AI simulation, modelling, and creating LLM or foundational models.

- Large-scale projects are based on **an open science approach, with a gradation in openness**. Earth-2 will offer a wide range of possibilities but uses Nivida internal tools. NASA and IBM are relying on Huggins Faces. NASA's first open-source geospatial artificial intelligence (AI) foundation model for Earth observation data model has many potential applications, including tracking changes in land use, monitoring natural disasters, and predicting crop yields. The HLS Geospatial FM is available at Hugging Face, a public repository for open-source machine learning models for the community. For Destination Earth, the DestinE Core Service Platform integrates and operates an open ecosystem of services (also referred to as the DESP Framework) to support DestinE-data exploitation and information sharing for the benefit of DestinE users and Third-Party entities.

- What distinguishes AI 4 Copernicus from the other platforms or projects is i) **a user-centric approach to develop a package of services** aiming to foster the development of AI services powered by EO in four strategic areas by companies (a bit like an AI/EO open innovation platform[20]), ii) an attempt **to reinforce a user-driven approach in searching and finding valuable and usable data** (semantic search, analysis-ready data in particular); the **integration of ethics in the service development**, which is key to provide explainability; iv) the recent focus on **education,** which is also a considerable challenge to address. Of Course, AI4 Copernicus is not the only project to implement such services. ESA is doing this (Cassini funds), and the AIODP also did it. It **provides a valuable direction for creating an AI node as the other platforms are developing powerful tools, but it does not entirely cover the AI4 Copernicus approach.**

- The objective of **creating an AI or an EO platform as a core node** is likely an illusion. The market is already buoyant and offers plenty of resources with a critical mass. The Huggins Face open AI model platform already has a $ 4.5M billion capitalisation. Stability AI, another startup proposing a platform with open, generative AI models, gathers a community of 200,000 developers. However, as Grega Milcinski, CEO of Sinergise, said in the Terrawatch podcast: “On the platform concept, I don’t believe in **one platform ruling them all**. Instead, let’s build Apis so users can tap into what they really need, bind them and develop services”. Federation between platforms is thus a better approach, as implemented today by the AIODP project.

[20] Open innovation platform organizing calls for solving challenges like wazokucrowd.com former innocentive

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 101016798.

### 4.2.2 The vision of the AI4 Copernicus partners

We interviewed partners to identify the priorities for an AI/EO node in an AI 4 Copernicus Exploitation Working group. After the seminar, we added a specific questionnaire to see where partners can work on potential features of the AI/EO node.

The table below summarises the results.

| FEATURES | Top ranked | Comments/Requirements |
|---|---|---|
| Access to EO data set via Diases and their services | 90 | Analysis ready data (preprocess to automatically curated datasets) for delivery to end-users |
| Storage and computing | 81 | Being able to run processes smoothly in any infrastructure. that could have an added value because of its capacity, access to data, scalability... |
| AI/EO library of use cases | 81 | The use case should be analysed by providing the context, requirements, and the best practices to solve it using AI/EO technologies |
| AI/EO repository (tools/algorithms) | 81 | Algorithms that could be reused in different workflows and could serve to homogenise processes and results, datasets. for training and tools for benchmarking. Different algorithms/tools should be intercompared. |
| Testing/Experimenting AI4 Experiment | 81 | AI4Experiment/CREODIAS integration will be an easy interface between data and output, allowing the user to develop pipelines efficiently and smartly. |
| Research and innovation challenges/projects | 72 | Possibly facilitating the creation of a "closed space" for industrial challenge-driven projects would be of value |
| Smart Aggregators | 65 | Additional smart aggregators as 3rd parties /intermediaries for AI training on demand, etc |
| Ethical Trustworthy AI (11) | 63 | Building in open-source code has to be automated with pool of providers of ethical AI assessment (third-party providers) |

*Figure 22 Most important features for an AI/EO node. Analysis from the AI4Copernicus partners*

The most important features relate to EO data (analysis-ready data) and the processing capabilities needed to develop AI models or algorithms. Those features are starting to be built with many initiatives and projects (see section before). Interestingly, two other features are ranked among the

priorities: the library of use cases and the Smart Aggregators[21]. They demonstrate the need for specific actions to mainstream EO: the first feature allows users to quickly understand what the added value of EO in their domain (showing by example) could be. The second feature demonstrates the need for specific intermediaries capable of bridging the gap between the techno-push approach predominant in the EO world and the solving problem approach required by average users. Indeed, the Job-To-Be Done approach is critical to success; what should Job the EO do in my domain?
Last interesting learning: the proposal to support research and innovation projects with a closed model (allowing industrials to share the results privately) with an open model where the IPR is open source. A balance between the two models should be found if a solid support for research and innovation projects is decided for an AI/EO node.

| FEATURES | Top ranked | Comments/Requirements |
|---|---|---|
| Matchmaking | 27 | There should be an automatic tool for this.<br>Nice, but not necessary again. It should be happening outside (we cannot build a tool that does everything) |
| Marketplace | 45 | Explore the EO market needs in relation to the current experimentation testing environment in the AIOD platform. |
| Search | 50 | A filter can be added rather than a dedicated search engine showing only the EO linked applications. Having a dedicated search engine seems too much. I think a dedicated search engine for the images available has more value than a search engine focused for the architectures.<br>See the EarthQA engine we have defined |
| General information about AI and EO | 54 | Combine information in a useful way.<br>2. By-products: events, newsletters, substack approach compiling or connecting to space publishing |

*Figure 23 Less important features for an AI/EO node. Analysis from the AI4Copernicus partner*

[21] TerraWatch definition. "Smart Aggregators are vertical-focused aggregators of relevant data, applying geospatial solutions to a specific market use case. Today, most of these aggregators have vertical-specific expertise and vertical-focused partnerships to source the right data for the analytics problem to be solved from upstream DaaS providers. We are seeing some very potent smart aggregators bringing together data operations expertise and machine learning prowess to solve unique challenges in specific verticals like precision agriculture, energy and utilities and mapping. – These companies are building large turn-key geospatial systems to solve specific problems.

The less critical features show less attention to commercial purposes such as the marketplace or matchmaking (here focused on competencies). Regarding the search, confusion happened between building a search engine from scratch and improving the search.

### 4.2.3 The end users' standpoints

The AI4 Copernicus AI Ecosystem Forum (Athens 28- 30 June) organised a workshop gathering 23 awarded consortia selected through a thorough open call process. Those consortia mainly comprise innovative SMEs, startups, and sometimes end-users. They will develop innovative AI services or products using EO Data and services provided by the AI4 Copernicus consortium in four areas: energy, security, agriculture, and health.
The workshop's objective was to collect their feedback on three main aspects: access to EO data, feedback on AI4 Copernicus services, and the following steps to foster more AI innovation using EO. The consortia were divided into three groups, corresponding to the different generations of open calls. The first group gathered the 1st and 4th open calls, the second group the 4th open call and the third group the 5th open call.
The three world cafes were managed by CF (topic 1 Access to EO data and services), the second by SATCEN (added value of AI4 Copernicus service) and third by BLS (A future? AIOD EXISTING & NEW SERVICES & AI/EO node).
However, the feedback collected from the OC winners was broader than only discussing the features of an AI/EO Node. Moreover, it represents more of the companies' user journey through AI and EO. It differs from the partners who could discuss precise features and rank them. Here, it's more the collection of needs and the different Job to Be Done that an AI/EO Node could develop. The next section will propose different scenarios that can address them.

#### *4.2.3.1 Feedback from the OC winners*

We summarise the main conclusions by highlighting ten messages.

1. **Filling a European gap at an affordable cost for Startups, Researchers and SMEs**

**Analysis-ready data** (ARD) is a growing demand in the EO world. ARD are preprocessed for users and 'ready to analyse'. They are typically made available as a data cube, specialised in a specific region or a thematic application. There's a growing demand for mainstreaming EO and AI as end-users who are not EO data experts **must find valuable and usable data** (see, for instance, Planetary variables or EarthBlox[22] low code product.)

The gap for startups and SMEs from our panel **is to access ARD at an affordable cost** to train their models. The market is in short supply, both for accessing the ARD with free or credit coins and enough High-Quality ARD. The link with the Earth Observation Training Data Lab (EOTDL) launched by ESA will be essential.

[22] The role of low-code Earth Observation tools in accelerating the sustainability transition

2. **Look for better and enhanced data fusion (combining satellite data with other data and internal workflows).**

The second data aspect raised by our panel is to focus on data fusion, including other data sources (in situ) and internal company workflows. A standard format to integrate EO, aerial, drones and economic social data could be explored. Beyond these two first messages, there is a need to build enabling tools so users can exploit EO data.

3. **Trustability of the data is an essential lever to create confidence in the market.**

Trustability in AI applications or models is critical to unleashing the market. It encompasses a wide range of activities, starting from data certification (can I trust the images, for instance?), explainability of the models (why your model is saying my home should not be insured?), and standards protocols for evaluation, benchmarks between the models to evaluate the performance and certify them (third part certification compliance). We are on a very sensitive line between innovation and regulation as startups frequently don't have a precise vision of their business models and products (too stringent regulation can undermine their efforts). But trust in the market will go through a set of tools described above that should not represent an insurmountable obstacle.

4. **"End-to-end environment for AI, data and experiments to create AI products.**

The question of the platform (AIODP) was constantly in the background of the discussions. There was a consensus to develop an end-to-end environment that could help startups and SMEs to use EO to create AI services or products, encompassing an extensive range of activities: ARD, computing, innovation services, experimentation etc.

The critical question is what and how. On what,

- Is it a HugginFace for satellite data with a library of certified and benchmarkable models/algorithms/pre-trained models? And open science project such as the NASA Foundational model on EO?
- Or a federation of platforms instead of one platform ruling all?

- Or training large amounts of data (who is doing this in Europe?)

5. **Super-resolution algorithms or high-resolution images** are needed for many applications where Copernicus resolution is insufficient, such as managing solar panels addressing heat urban islands, etc. Private suppliers can provide the resolution, but as mentioned above, Startups and SMEs may not be able to afford these services when developing an innovative solution.

6. **Mainstreaming EO and scaling up the market will require supply and demand-driven actions.**

Mainstreaming the market requires enabling actions that convince thousands of users to use EO daily! It's easier to say than doing it in a very tech-driven industry that is EO. That will likely require a mental shift in this industry to focus on solving user's problems rather than providing data,

computation power, etc., which are still essential but should be in the shadow. SPAAS (Solving Problems As A Service) is undoubtedly a promising route illustrated by pioneers such as Sinergise, Planetary Variables or Earth Blox. Our panel of consortia has issued proposals to foster this approach.

- Providing package services to analyse, train data and models.
- Incentivise more aggressively end-users to invest in satellite data, even enforce the use of EO data in public procurement, for instance.
- Develop third parties/Brokers capable of balancing expectations towards EO data and the real needs from an end-user perspective (more market-oriented than techno push). To build the market for low-tech users, there is a need to have brokers that can bridge the gap between the demand side (user side) and the supply side to safeguard interests from the user standpoint and search for specific expertise. This intermediary layer is shown in the diagram below as a critical condition for mainstreaming AI and shifting from a market composed of innovators to a market with an early majority.

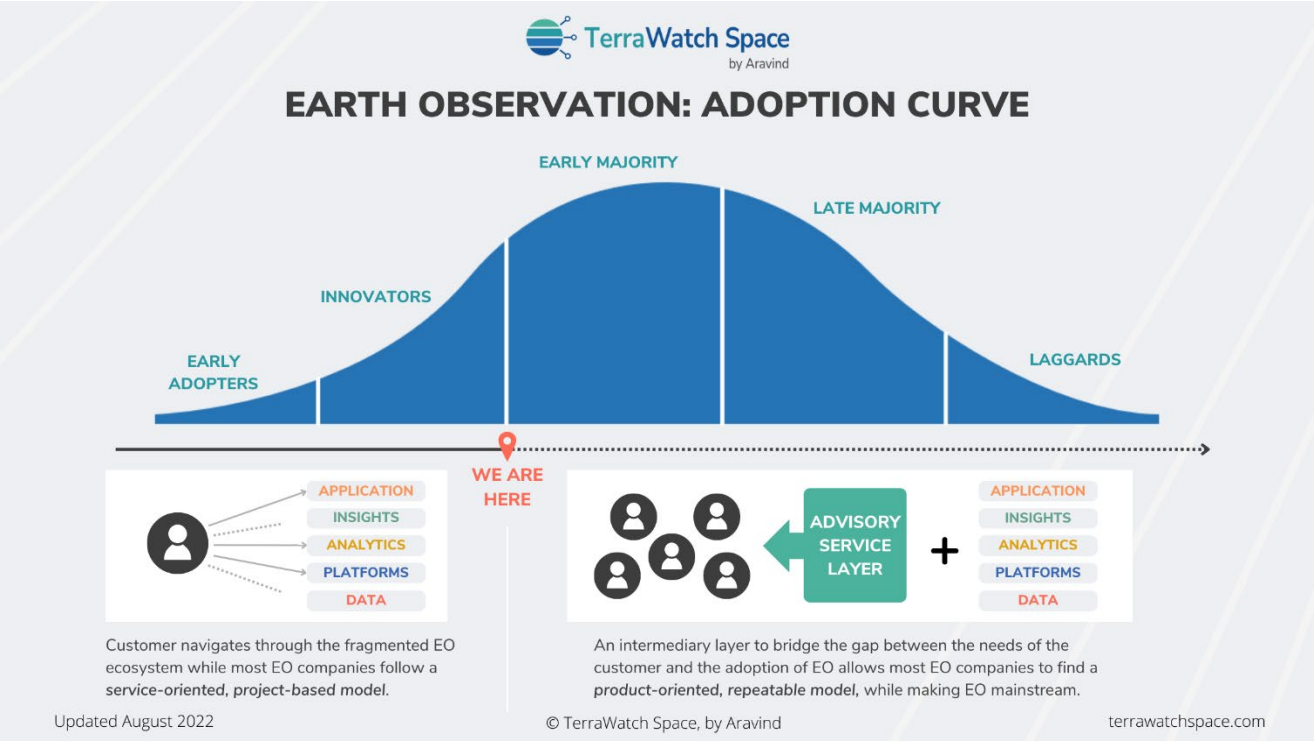



*Figure 24 Earth Observation Curve Terra Watch*

**7. Mainstreaming EO and scaling up the market is also a question of Return On Investment. How to make money and turn upside down the current tech push reasoning to a solving user problem approach first!**

It's difficult for a startup developing downstream applications to make money. It seems the companies in a better position are tech companies delivering services such as data, infrastructures, and computing. Balancing this market power is vital as downstream applications are essential to solve critical challenges (climate change, agriculture, energy etc.) Proposals were made by our panel of consortia.

- Launch more large-scale projects on several verticals aiming to trigger the market with large companies or end-users (raising challenges) and startups/SMEs/research developing using satellite data to solve problems at scale. (for instance, building a Generative AI model on climate change with insurers and banks?). The proposal was close to what ESA is doing with its SCOUT program[23].
- Develop marketplaces to create visibility and potential revenues.

| Getting access to useful and usable data is not so easy for non « usual suspects » | AI ML models : transparency, open source and explainability | The industry is tech driven not demand side |
|---|---|---|
| Users are from different worlds and often don't understand EO data, and EO data are in competition in budget allocation with more controlled sources of data or more usual spending<br><br>Users can tend to overestimate what they need (high resolution where maybe it is not needed)<br><br>EO data must be easily accessible and embedded in company workflows to create value | Many companies are developing predictive ML Models that, for example, can assess climate risks for households or industrial assets.<br><br>Benchmarking and explainability are needed to explain the decision<br>« I can't no longer insure your house because my model is telling me not. »<br><br>Computing power, and data sets are necessary to train models | EO is tech push, not necessarily organised around solving user problems<br><br>Documentation, and access to assets and repositories are essential, but there is a need to see brokers arise to fill the gap between both sides<br>• User<br>• EO/AI industry<br><br>Showing concrete examples is important as frequently the IT department needs to convince the C suite or the founders. Especially in areas where the decision for budget allocation can have an impact on human life (humanitarian, health |

*Figure 25 Three challenges to mainstream AI products or services powered by EO BLS*

**8. EO Community is fragmented and requires more interdisciplinary education and research programs.**

Developing AI products or services by using EO requires multidisciplinary skills and competencies. Bringing EO expertise, mathematics, physics, IT, and vertical (of a specific domain such as agriculture, climate change, etc.) expertise is crucial. Thus, the talent shortage in Europe needs to be tackled. To that end, educational courses and research programs should develop relevant interdisciplinary programs. Education programs should also target end-users and SMEs, foster open collaboration, create new services for companies such as -» Pitch my needs" to other SMEs for cooperation, find experts (experts-as-a-service), data, etc. (i.e., chatbot service)

**9. IPR issues need to deal with a new situation related to the ownership of datasets incorporated in a trained model and reused.**

**10. Generative AI will likely invent EO Large Langage Models.**

We have seen groundbreaking progress in Large Language Models (AI) and Imagery (EO) in recent years. LLM or generative AI will likely be specialised in the following years, allowing users to use EO data quickly and efficiently. We can see already promising and pioneering projects in this field with the NASA/IBM/Hugging Faces open foundational models or the recent publication of a research

[23] https://www.esa.int/Applications/Observing_the_Earth/FutureEO/Scouts_ESA_s_agile_research_missions

paper, EarthPT: a foundation model for Earth Observation[24]. The critical question is the capacity of Europe to build such an open foundational model, relying also on open sciences or project approach to develop the model.

## 4.3 Pathways for an AI/EO node on the AIODP

### 4.3.1 AI/EO Node Building blocks

The diagram next page shows a potential pathway to develop an AI/EO node that could represent a bridge between EO and AI. It would be a federation of different platforms and projects while developing a set of services in the continuity of the AI4 Copernicus project.

The maps show the AI/EO node of the AIODP **as a bridge between EO and AI** to develop more downstream applications, models, products, and services. We can find in orange essential services developed by ending projects or other providers (private providers, such as Earth Blox or Hugging Face). We selected the most important one regarding their synergy potential.

**On the left side, EO-powered projects** cover critical activities to make EO data accessible, valuable, and helpful, ready to be transformed into applications or insights, such as Copernicus Data Ecosystem, Europe DATA Cube and EOTDL plus large-scale and outstanding projects like Destination Earth.
It is worth noting that some private providers are trying to develop affordable access to ARD. through low-code platforms such as Earth Blox or packaging datasets in ready-to-use products (planetary variables from Planet).
"*Thanks to the game-changing rise of serverless native cloud infrastructures, a fourth approach is emerging in the EO industry. Low-code EO platforms now enable the creation of customised EO insights at scale. These cloud web-based tools allow processing and analysing EO data in no time to derive actionable insights. Some applications require no coding (for simple applications), while others require a limited amount of coding (for the most advanced, less mainstream applications).*

*In the case of EO, low-code EO platforms present the additional advantage of removing the need for a deep understanding of the peculiarities of remote sensor types and design, e.g. orbits, bands, repeat periods and knowledge of how the electromagnetic signals interact with different objects and surfaces. These now enable both the EO experts and non-experts to focus on the business objectives rather than on the mechanics of remote sensing."*[25]

[24] EarthPT: a foundation model for Earth Observation Michael J. Smith∗ Luke Fleming James E. Geach Aspia Space Ltd., Cornwall, UK
[25] The role of low-code Earth Observation tools in accelerating the sustainability transition. Terra Watch Earth Blox 2023

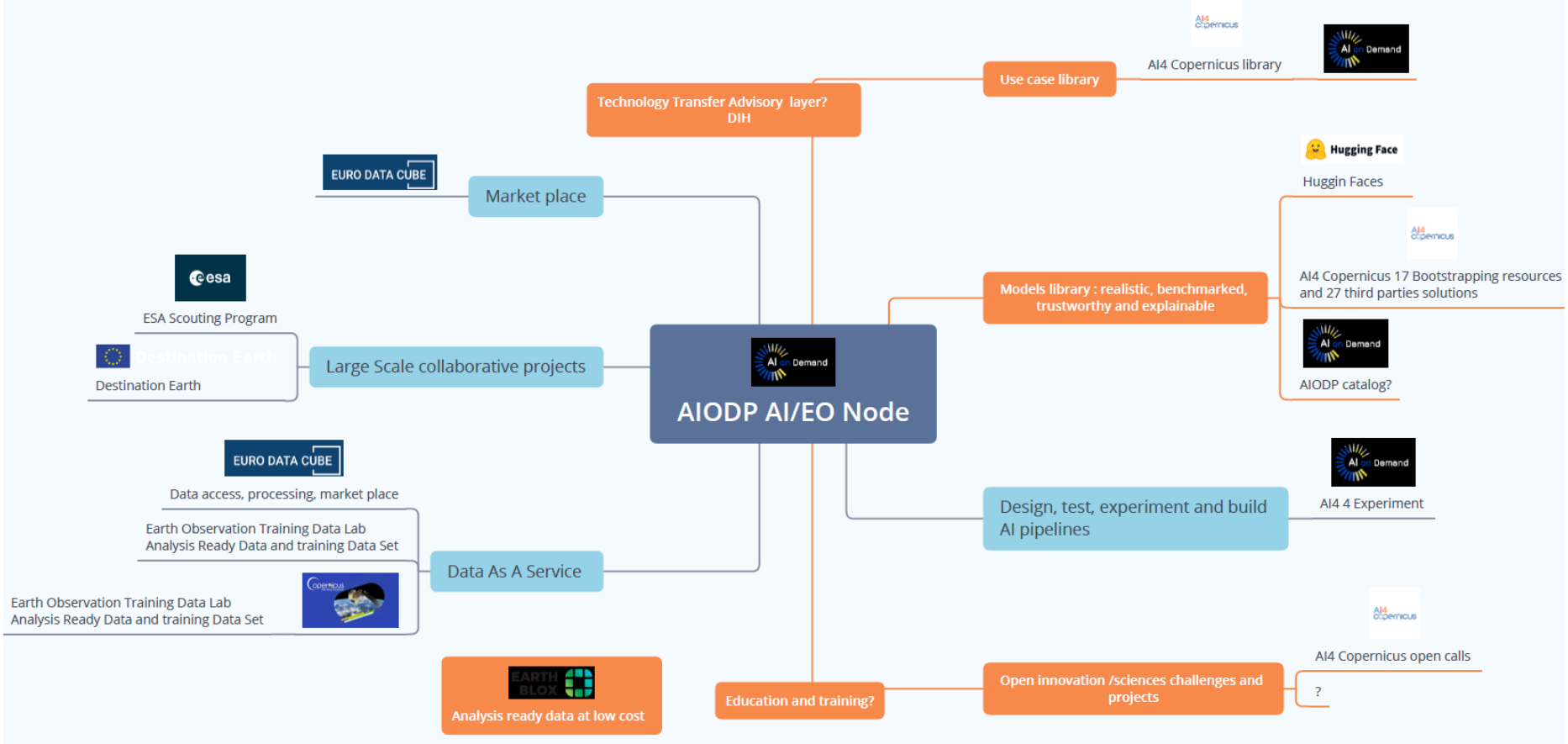



*Figure 26 An AI/EO node as a federation of platforms while developing services*

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 101016798.

On the right side, **we can find services and features expected to mainstream** the usage of AI and EO. The AIODP is developing services around the AI4 Experiment or the catalogue. AI4 Copernicus has deployed some in orange, and the results are now hosted in the AIODP. Those services must be improved and updated and reach a critical mass to attract users and mainstream usage.

- **The use case library** intends to **raise users' awareness** of what can be done with EO data combined with AI. Its user stories and qualified information. AI4 Copernicus provided a first set of user stories through the open calls' winners. It's a first step.
- **A library of open, explainable, certified models with benchmarks** could represent an important step to give confidence to the market and simultaneously trigger positive loops to develop the models (see Huggins Face, Stability AI open projects for a benchmark). On the first aspect, explainability and benchmarking of the performance of the models or algorithms will become essential, particularly in the climate change solutions fields (how to explain decisions taken by the algorithm for insurance, for example)

**Open innovation and challenges are a third critical aspect.** ESA partially covers this through the Cassini fund or the SCOUT program for dedicated activities. However, there are a lot of expectations from the open-call winners to find support not only **in the innovation stage but also in creating the conditions to scale their innovation on the market. Several possibilities could be explored.** Firstly, create the conditions to attract large companies or consortiums to set challenges and work with innovative companies to solve a problem at a one billion market minimum. Secondly, exploring a duplication of the Scout program.

The gap between the current state of the art (GPT-4) and open-source models is changing quickly regarding open science. The open-source community has solved many scaling problems through a range of optimisations. MosaicML is an excellent example, demonstrating that they can train a Stable Diffusion model, which is not a large language model, six times cheaper than the original. The dynamic of the open-source market creates a faster rate of iteration. Potentially, open-source models could learn and iterate far faster than closed-source ones[26].

**In the middle of the map**, we can find two essential building blocks as they address critical barriers.

- The first one is the **advisory layer** able to bridge the gap between, on the one hand, the complexity of EO data and, on the other hand, the problem to solve affordably and efficiently.
- The second one is to address **the competencies gap and shortage**. Developing AI applications or solutions requires a high level of competencies and interdisciplinary which are both in short supply.

### 4.3.2 AI/EO node potential models and targets

It seems challenging to cover all the targets and develop all the building blocks. Different pathways can be taken, and models selected between

26 Azem Azhar The open-source future of artificial intelligence May 2023

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 101016798.

- A proposal **for well-experienced users of EO to facilitate and speed up the development of EO/AI applications and services**. The platform could be like an AI/EO Innocentive model with links to the EODTL, Copernicus Data Ecosystem and Europe Data Cube and additional services such as analytics and insight (how to use EO). https://www.innocentive.com/ . A similar program to SCOUT opened to large industrial companies to develop jointly with startups and SMEs 1 billion Euros market application could be explored. The model could provide open and close results. A library of open, certified and explainable models should be built.

- A **second model could target technology transfer and algorithm engineering** by providing services detailed in the diagram below. The model could totally or partially cover all the steps needed to develop algorithms.

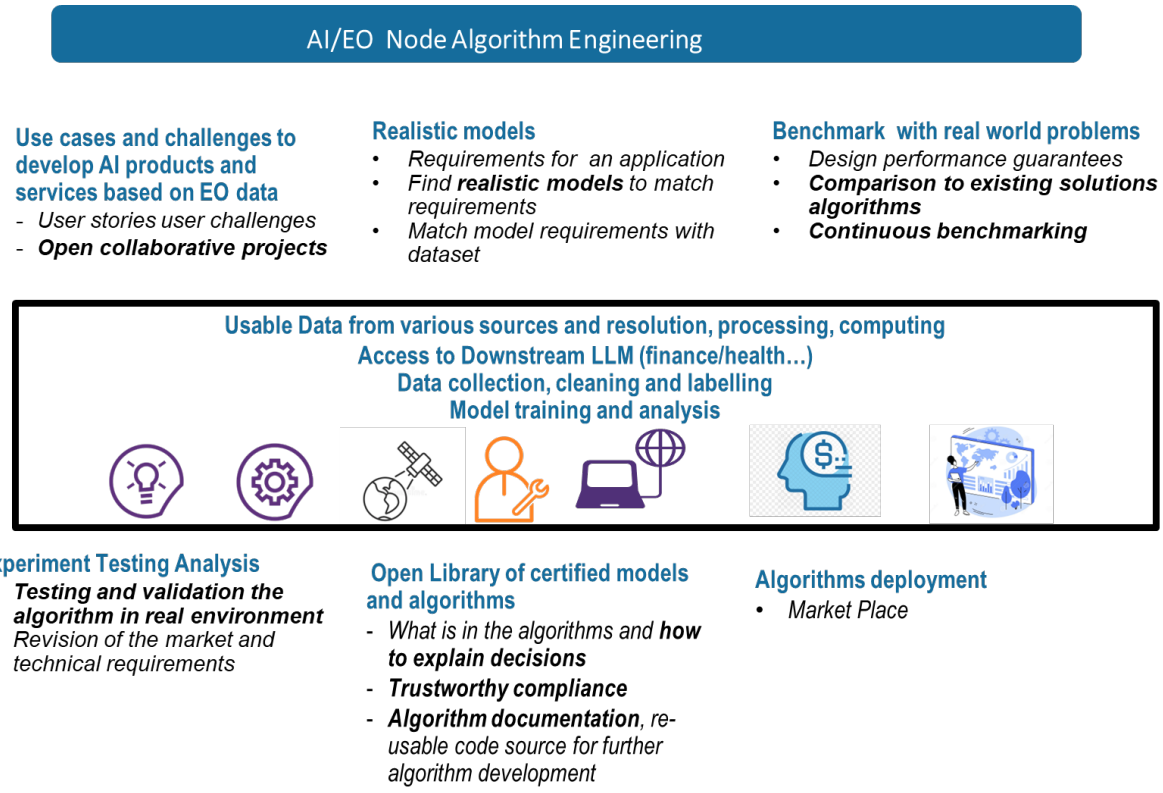



*Figure 27 An AI/EO node focusing on technology transfer and algorithm engineering along the value chain*

The third one **could target low-end tech users to advocate for the use and provide services that can help users define what they want to do,** how it can be done, what resources to commit and for what expected results outcomes. The focus is on the ARD and Advisory Layer that DIH can represent. User stories library and the AIODP catalog should progressively reach a critical mass. The focus will be on technology transfer and algorithm engineering.

- The fourth possibility is to **focus on research and education around large foundational models for EO or strategic research programs** that can be built using a federated platform approach, AI4 Experiment and the library of open models' deployment. Self-driving labs coupled with open science projects could be a way to develop programs. Education components will have to be considerably strengthened.

The new project for the AIOD, Deploy AI, would likely invest in these different models to find the best or the best combination.

# 5 Conclusion

AI4 Copernicus delivered impressive results and paved the way tro transition from a techno-push EO market to a user-centric approach, enabling faster deployment of AI solutions. Mainstreaming AI/EO will allow to address critical challenges. This will also require completing the start-ups and SMEs support by including scaling up by design support combining ethical, business and financing supports across the different stages and Valley of Death.
Fostering a more innovative ecosystem is scattered across many projects today, all deploying exciting and added value. The follow-up Deploy AI project could develop an AI/EO node with different scenarios or combinations between them:i) **facilitate and speed up the development of EO/AI applications and services**; ii) **technology transfer and algorithm engineering; iii) could target low-end tech users; iv) focus on research and education around large foundational models for EO or strategic research programs.**

# 6 Annex

## 6.1 Lessons learned in each of the world's cafes

The tables below show the results of the three world cafes.

**World café 1 Access to data**

### Highest Priority (High Importance-High Impact)

**Data Group A (1st & 4th OC)**

**HIGHEST PRIORITY**

Analysis ready data

**MEDIUM PRIORITY**

low latency
fast query as big query
more learning resources
tutorials & docs
compatibility with cloud platform

**Data Group B (3rd OC)**

**HIGHEST PRIORITY**

data availability
MS data with higher temporal resolution

**MEDIUM PRIORITY**

Access to socioeconomic data
Visualize spatiotemporal data
Federated cloud
Preprocess and regrid data

**Data Group (5th OC)**

**HIGHEST PRIORITY**

VHR Data availability
CAMS data
Expand data catalog

**MEDIUM PRIORITY**

Examples & Docs
Access to data generated by AI4Cop
Combine user data with EO data (Data fusion)

### Key discussion topics

**Data Group A (1st & 4th OC)**

1. standard access to data (APIs, etc)
2. standard formats to support cloud environments
3. tutorials and documentations
4. fast query and download with low latency
5. data availability
6. analysis ready data

**Data Group B (3rd OC)**

1. data availability
2. preprocessing and regridding
3. visualization of spatiotemporal data through user friendly GUIs

**Data Group (5th OC)**

1. analysis ready data
2. data availability from EO, CAMS archives, VHR data
3. data preprocessing and fusion
4. examples and docs

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 101016798.

**World Café 2 AI4 Copernicus Services**

## Highest Priority (High Importance-High Impact)

### Group A (1st & 4th OC)

**HIGH PRIORITY**

Step by step documentation and tutorials for not technical people

National language guidance for high tech users

**MEDIUM PRIORITY**

Documentation available for low tech users to implement

As a user I want access to higher level services

Processing services saves as libraries in Py, C++

### Group B (3rd OC)

**HIGH PRIORITY**

Pretrained models

Obtain S2 data with higher temporal resolution

**MEDIUM PRIORITY**

New multispectral data by merging constellations

Easy tiles collection

Creation of mosaic timelines

High quality oceanography data sets suitable for all areas of the globe

### Group (5th OC)

**HIGH PRIORITY**

High resolution of the input

**MEDIUM PRIORITY**

Integrate semantic resolution algorithms

DEM generation from c-band and l-band sentinel 1

## Key discussion topics

### Group A (1st & 4th OC)

Easy integration of services

Clear information in guidelines

Hand to integrate onto the already existing infrastructure

Efficient for heavy processing

SNAP is pretty slow compared to alternatives

Not good API access for S2L1C AOI data

Not so easy to use for not technical people

### Group B (3rd OC)

Bigger tutorial

A service to provide the mosaic of the tiles in a timeline

Updated and denser pre-trained models

More data for training; eg for instance crop data

### Group (5th OC)

Better ways to request more specific data

Integration of bootstrapping services and AIOD more documentation

Land cover available products

Use semantic data to integrate easily data sources

Linked data tool

Integration of multiscale data

**World Cafe 3 AI/EO node**

## PRIORITIES

### GROUP A + 5th HOC
### HIGH PRIORITY
Build big pilot use-cases that you subsidise funds to be able to communicate results
Awareness creation
Economic sustainability for SMEs (ability to stay financially alive)
Educate the end-user (NGO, public sector, organisations)

### MEDIUM PRIORITY
End-user engagement – understand their needs
3rd party independent validation of AI results

### GROUP B = 3rd OC
### HIGH PRIORITY
Marketplace for data
Trainings, tutorials, building network of learners

### MEDIUM PRIORITY
Repository of quality, tested controlled pre-trained models
- Adaptable or additional datasets for innovative models, services
- Foster collaboration between private companies to create faster new solutions
- "Pitch my need" (experts,
- 3rd party as intermediary to foster collaboration with SMEs in the EO domain
- research consortium, companies)

### GROUP C = 1st & 4th OC
### HIGH PRIORITY
Free periods to test services, tools, etc
Large scale pilots to create the conditions for scaling a market

### MEDIUM PRIORITY
Certified and trustworthy data for start-ups
End-to-end integrated environment for AI,data, experiments

## Key Identified Needs (5th OC)

Need 1: Independent validation framework
Need 2: Test the algorithms on the other case studies
Need 3: Need for real problem use cases from clients (market-driven not technology-driven) /need for more end-user focus to build demand
Need 4: Ability to share easily (between community members)
Need 5: Data ownership know-how, IP and data ownership protection and security for data, models and algorithms
Need 6: Access to certified data, good quality data
Need 7: Use the algorithms in an easy way – easy to use training models (ability to test them for free)
Need 8: Provide guidelines/specifications for non-tech people (collaboration between interdisciplinary companies (within AI4Copernicus and external ones)
Need 9: knowledge sharing for data, benchmarks by academic partners (uniformity in benchmarks)
Need 10: gather resources so as to work as a DAO
Need 11: need for economic sustainability for SMEs in AIEO
Need 12: compliance services
Need 13: Data harmonisation – need for good AI/data standards & a repository for registered good AI standards
Need 14: Need for education on various areas (specialized AI-EO and business, legal (IPR, etc) issues)

# Key Identified Needs (3rd OC)

Need 1: Funding needs & public funding
Need 2: Library of typical models for classic tasks (superresolutions, forecasting with pre-trained model weights)
Need 3: Need to standardised datasets
Need 4: A new service - "Pitch my needs" to other SMEs for collaboration, to find experts (experts-as-a-service), data, etc (i.e., chatbot service)
Need 5: Training tutorials and need to meet other learners
Need 6: Access to annotated data
Need 7: Sharing expertise (privately), maybe a 3rd party can ensure, guarantee security
Need 8: Repository of pre-trained models with benchmarks that ensure quality
Need 9: Marketplace for data-assets
Need 10: Ecosystem with connected labelled data and user stories
Need 11: Need to interconnect models and data (i.e., hydrological models and data) via an easy interface
Need 12:Need for collaboration between SMEs (primarily and then with others like DIHs, etc)
Need 13: Need for expertise sharing (privately) via a 3rd party to guarantee security
Need 14: Repository of pre-trained models with benchmarks (freemium model: free models and fee-based higher quality models)
Need 15: Open library of certified models
Need 16: Benchmarks of data to be provided by Copernicus with "nice" documentation
Need 17: Marketplace for data assets (buy and sell datasets)
Need 18: Standard datasets for evaluation purposes

# Key Identified Needs (1st & 4th OC)

Need 1: Repository with AIEO challenges for different domains
Need 2: Open libraries to develop custom use cases
Need 3: A use-cases catalogue from different domains
Need 4: Access to high-quality data and a way to load them easily, before starting an analysis
Need 5: Harmonised (specialized) data-sets (i.e., hydrological datasets, etc) and facilitate data that are in the same format
Need 6: Repository of registered providers of good AI and data providers
Need 7: Ensure a trustworthy AIOD platform & AIEO Node (IPR issues when using private data or algorithms, Security, standardized, compliant AI and data, etc)
Need 8: Different tiers of services that is time-dependent (i.e., Tier 1: 2 months free (for a testing period), Tier 2: fee-based)
Need 9: Tailored education (i.e., Exploitation aspects of OSS services, etc)
Need 10: An integrated environment – that can be used as as service
Need 11: Collaboration tools to integrate, acknowledge and results
Need 12: Easy to use training models, modules

## 6.2 Assets developed by OC consortiums

| Project | AI assets |
|---|---|
| SR4C3 | https://www.ai4europe.eu/research/ai-catalog/environment-monitoring-ukraine-conflict-areas<br>https://www.ai4europe.eu/research/ai-catalog/environment-monitoring-mali-conflict-areas |
| SCAVIHO | https://www.ai4europe.eu/research/ai-catalog/dataset-phenological-stages-grapevines<br>https://www.ai4europe.eu/research/ai-catalog/automated-pipeline-real-time-map-generation-and-scalable-nvdi-values |
| SLIDE | https://www.ai4europe.eu/research/ai-catalog/slide-satellite-images-prediction-deep-learning |
| Sen4Weeds | https://www.ai4europe.eu/research/ai-catalog/automatic-detection-field-weeds-using-ai |
| VALENS | https://www.ai4europe.eu/research/ai-catalog/vake-pathfinder<br>https://www.ai4europe.eu/research/ai-catalog/vake-overwatch<br>https://www.ai4europe.eu/research/ai-catalog/rf-ais-dashboard |
| HumanityWatch | https://www.ai4europe.eu/research/ai-catalog/quanteo-0 |
| ODFuse4Ship | https://www.ai4europe.eu/research/ai-catalog/hires-current-med-2022 |
| LobeliaAir | https://www.ai4europe.eu/research/ai-catalog/sensor-community-ai-calibrated-data-sofia-jan-may-2022 |
| EO4NOWCAST | https://www.ai4europe.eu/research/ai-catalog/eo4nowcast-near-real-time-soil-moisture-assessment-and-pluvial-flood-nowcasting<br>https://www.ai4europe.eu/research/ai-catalog/eo4nowcast-near-real-time-soil-moisture-assessment-and-pluvial-flood-0<br>https://www.ai4europe.eu/research/ai-catalog/eo4nowcast-near-real-time-soil-moisture-assessment-and-pluvial-flood-1 |
| ESFA | https://www.ai4europe.eu/research/ai-catalog/geoskop-esfa-time-series-forecaster-component<br>https://www.ai4europe.eu/research/ai-catalog/geoskop-esfa-time-series-forecaster-component |
| PLANET | https://www.ai4europe.eu/research/ai-catalog/planet-hyper-local-climate-driven-tool |
| FertiRec | https://www.ai4europe.eu/research/ai-catalog/postcode-based-fertilizer-rate-recommendation-system |
| OPTIMAL | https://www.ai4europe.eu/research/ai-catalog/optimal-copernicus-irrigation-management-toolkit-environmental-parameters |
| LIVE4ENV | https://www.ai4europe.eu/research/ai-catalog/labels-land-cover-classification-livestock-farms |
| Urbalytics | https://www.ai4europe.eu/research/ai-catalog/urbalytics |
| AI4EO.Green | https://www.ai4europe.eu/research/ai-catalog/ai4-e2ogreen-ai-powered-pipeline-golf-and-grassland-monitoring |
| AIRON MAN | https://www.ai4europe.eu/research/ai-catalog/ai-ron-man-wildfire-hazard-risk-assessment-backend-service<br>https://www.ai4europe.eu/research/ai-catalog/ai-ron-man-wildfire-hazard-risk-assessment-webapp<br>https://www.ai4europe.eu/research/ai-catalog/ai-ron-man-wildfire-hazard-risk-assessment-pipeline |